

Univerzita Karlova v Praze  
Filozofická fakulta  
Ústav českého jazyka a teorie komunikace

Diplomová práce

Jiří Lebeda

# **Morf·em·at·ic·k·á strukt·ur·a sou·čas·n·é češ·t·in·y**

Od lingvistické teorie k automatickému počítačovému zpracování

Morphemic Structure of Contemporary Czech  
From linguistic theory to automatic computer processing

Vedoucí práce: doc. RNDr. Vladimír Petkevič, CSc.

Praha 2010

## **Prohlášení**

Prohlašuji, že jsem diplomovou práci vypracoval samostatně s využitím uvedených pramenů a literatury.



.....  
Jiří Lebeda

V Praze dne 31. 8. 2010

## **Poděkování**

Rodičům za podporu, nepodmiňovanou ani v nejmenším nutností porozumět všem tématům, jimž jsem se v průběhu svého úředně dlouhého studia věnoval.

Vedoucímu práce, doc. Vladimíru Petkevičovi, za citlivý, „minimálněintervenci“ přístup.

Doc. Karlu Olivovi za míru, v jaké při svém působení ve funkci ředitele ÚJČ AV ČR navázal na zářnou ústavní tradici morfematického výzkumu z šedesátých let 20. století, a za příkladně konsekventní postoj k možnostem rozvinout všestrannou spolupráci Ústavu s autorem těchto řádků. Jeho nevědomé zásluhy na nesnadném zrodu této práce zaujímají v dosud nenapsaných pamětech podstatnou část kapitoly věnované létům 2004 až 2010.

Pokládám rovněž za milou povinnost uvést, že vznik předkládané práce nepodpořil žádný český ani evropský grantový systém.

## OBSAH

<b>0</b>	<b>EXOLINGVÁLNÍ PROLEGOMENA .....</b>	<b>10</b>
0.1	JAZYKOVÁ INFRASTRUKTURA JAKO POJÍTKO MEZI TRADICÍ A INOVACÍ .....	10
0.2	BÁDÁNÍ NEZATÍŽENÉ STARŠÍMI PŘÍSTUPY K JAZYKU A TZV. BEZHODNOTOVÁ LINGVISTIKA .....	12
0.3	HOMO THEORETICUS A HOMO PRACTICUS.....	14
0.4	SYNERGIE, INTEROPERABILITA A MORFOLOGICKÉ KONKLÁVE.....	17
0.5	METODY, VÝCHODISKA A CÍLE TÉTO PRÁCE .....	18
<b>1</b>	<b>ANIZOTROPIE JAZYKOVÝCH JEDNOTEK V EMPIRICKÉM SVĚTĚ (SÉMOTICKÉ EXPOZÉ).....</b>	<b>21</b>
1.0	MOTTO .....	21
1.1	.....	21
1.2	.....	21
1.3	.....	22
1.4	.....	24
1.5	.....	26
1.6	.....	27
1.7	SHRNUTÍ.....	28
<b>2</b>	<b>MORFÉM JAKO KONSTITUENT MENTÁLNÍ REPREZENTACE MYSLI .....</b>	<b>29</b>
2.0	MOTTO .....	29
2.1	VÝCHODISKA.....	29
2.2	SOUVZTAŽNOST PSANÉ A MLUVENÉ FORMY JAZYKA .....	29
2.3	ORGANIZACE MENTÁLNÍHO SLOVNÍKU .....	30
2.4	MENTÁLNÍ SLOVNÍK A PRINCIPY JEHO FUNGOVÁNÍ.....	32
2.5	SHRNUTÍ.....	33
<b>3</b>	<b>POJEM MORFÉMU V DĚJINÁCH (SVĚTOVÉ A) ČESKÉ LINGVISTIKY .....</b>	<b>34</b>
3.1	MORFÉM JAKO TERMÍN (A POJEM) .....	34
3.2	KOMENSKÝ (REKONSTRUOVÁNÍ KOŘENOVÉ A JANUA) .....	36
3.3	ROSA (GRAMATIKA A SLOVNÍK).....	36
3.4	DOBROVSKÝ (ÜBER DEN URSPRUNG, BILDSAMKEIT A MLUVNICE) .....	37
3.5	ČELAKOVSKÝ (KARTOTÉKY A NEREALIZOVANÉ PLÁNY) .....	39
3.6	SHRNUTÍ.....	43
<b>4</b>	<b>MORFEMATICKÁ ROVINA A JEJÍ JEDNOTKY V SYSTÉMU JAZYKA.....</b>	<b>44</b>
4.0	.....	44
4.1	MORFÉM FLUKTUUJÍCÍ NAPŘÍČ ABSTRAKTNÍMI ROVINAMI POPISU.....	44
4.2	REPETITORIUM POZNATKŮ (MORFEMATIKA, FORMÁLNÍ MORFOLOGIE, MORFOTAKTIKA, MORFONOLOGIE) ..	46
4.3	MORFÉMIKA/MORFEMATIKA .....	46
4.3.1	Základní jednotky (nacionálně, třídění) .....	46
4.3.1.1	morfém .....	46
4.3.1.1.1	definice .....	46
4.3.1.1.2	třídění .....	46
4.3.1.2	submorfém = konektém {K}.....	46
4.3.1.2.1	definice .....	46
4.3.1.2.2	třídění .....	47
4.3.1.3	morf .....	47
4.3.1.3.1	definice .....	47
4.3.1.3.2	třídění .....	47
4.3.1.4	submorf = konekt {K} .....	47
4.3.1.4.1	definice .....	47
4.3.1.4.2	třídění .....	47

4.3.1.5 alomorf = varianta morfému {M1} .....	47
4.3.1.5.1 definice .....	47
4.3.1.6 izomorfa .....	48
4.3.1.6.1 definice .....	48
4.3.1.7 morfoném .....	48
4.3.1.7.1 definice .....	48
4.3.2 Morfologická paradigmatika .....	48
4.3.2.1 slovní tvar = morfologická forma {E} .....	48
4.3.2.1.1 složený/analytický/opisný/perifrastický tvar .....	48
4.3.2.2 morfologická báze (slovního tvaru) .....	49
4.3.2.3 slovotvorný základ .....	49
4.3.2.4 tvarotvorný základ .....	49
4.3.2.5 (tvaroslovný/tvarotvorný) kmen .....	49
4.3.2.5.1 jednoduchý × odvozený .....	49
4.3.2.5.2 infinitivní, přítomný, minulý .....	49
4.3.2.5.3 pasivní, imperativní .....	49
4.3.2.6 paradigma .....	49
4.3.2.6.1 kategoriální × formální .....	49
4.3.2.6.2 široké × úzké .....	49
4.3.2.6.3 morfologické paradigma (slovních druhů) .....	50
4.3.2.7 morfologické typy .....	50
4.3.2.8 slovotvorné typy .....	50
4.3.2.9 pád .....	50
4.3.2.10 pádová polymorfie .....	50
4.3.2.11 homomorfie .....	50
4.3.3 Funkční a konstrukční typologie morfémů .....	50
4.3.3.1 třídění .....	50
4.3.3.1.1 vázané × nevázané / volné .....	50
4.3.3.1.2 samostatné .....	50
4.3.3.1.3 funkční × prázdné .....	50
4.3.3.1.4 jádrový .....	50
4.3.3.1.5 hypermorfém (rozšířený) .....	50
4.3.3.1.6 bimorf .....	50
4.3.3.2 kořen(y) / radix(y) .....	50
4.3.3.2.1 reziduální, zbytkový .....	51
4.3.3.2.2 jaderné (a tzv. rozšiřující elementy) .....	51
4.3.3.2.3 reziduální .....	51
4.3.3.3 afixy .....	51
4.3.3.3.1 gramatické/tvaroslovné/tvarotvorné .....	51
4.3.3.3.1.1 definice .....	51
4.3.3.3.1.2 třídění .....	51
4.3.3.3.1.3 kumulativní morfémy .....	51
4.3.3.3.2 slovotvorné .....	52
4.3.3.3.3 prefixy/předpony .....	52
4.3.3.3.4 prefixoid .....	52
4.3.3.3.5 sufixy/přípony .....	52
4.3.3.3.5.1 finální/koncovka .....	52
4.3.3.3.5.2 sufixy funkční × prázdné .....	52
4.3.3.3.5.3 tzv. složené sufixální morfémy .....	52
4.3.3.3.5.3 spolumorfém (sufixu) .....	53
4.3.3.3.5.4 derivační .....	53
4.3.3.3.5.5 modifikační .....	53
4.3.3.3.6 sufixoid .....	53
4.3.3.3.7 postfixy .....	53
4.3.3.3.8 postfixoid .....	53
4.3.3.3.9 infixy a interfixy .....	53
4.3.3.3.9.1 infix .....	53
4.3.3.3.9.2 interfix .....	53

4.3.3.3.10 raritní afix .....	53
4.3.3.3.11 unikátní afix .....	53
4.3.3.4 afixoidy .....	53
4.3.3.5 lexikální morfémy .....	54
4.3.3.6 alternační morfémy .....	54
4.3.3.7 asociované (sdružené) morfémy = cirkumfixy {Č}, rozštěpené morfémy {Sl} .....	54
4.3.3.8 konektémy (konekt – alokonekt) = submorfémy {M1, So} .....	54
4.3.3.8.1 redundantní, prázdné morfémy .....	54
4.3.3.8.2 složený slootovorný morfém .....	54
4.3.3.8.3 kmenotvorný morfém .....	54
4.3.3.9 kmenový konektém / kmenotvorný morfém {E} .....	54
4.3.4 Identifikace alomorfů .....	54
4.3.4.1 alomorf (alokonekt) téhož morfému (podmínky): .....	54
4.3.4.2 typy alomorfů .....	55
4.3.4.2.1 nekomplementární .....	55
4.3.4.2.2 komplementární .....	55
4.3.4.2.2.1 paralelní .....	55
4.3.4.2.2.2 synonymní .....	55
4.3.4.2.2.3 supletivní .....	55
4.3.5 Souvztažné pojmy .....	55
4.3.5.1 nulový morfém (konektém) / morfémová nula {E} .....	55
4.3.5.2 morfémový/morfematický šev .....	56
4.3.5.3 morfematický / morfémový uzel .....	56
4.3.5.4 perintegrace / přerozdělení {E} .....	56
4.3.5.5 deetymologizace .....	56
4.3.5.6 *trunkace/trunkácia .....	56
4.3.5.7 distribuce (libovolného morfému) .....	56
4.3.5.8 centrum a periférie .....	56
4.3.5.8.1 centrum .....	56
4.3.5.8.2 periférie .....	56
4.3.5.9 morfostylém .....	56
4.3.5.10 odpovědnost a kompetence morfému .....	56
4.3.5.10.1 odpovědnost morfému .....	56
4.3.5.10.1 kompetence morfému .....	57
4.3.5.11 fúze .....	57
4.4 MORFOTAKTIKA .....	57
4.4.1 definice .....	57
4.3.2 obecné poznámky .....	57
4.3.3 fonematická struktury morfů .....	57
4.3.3.1 čeština .....	57
4.3.3.2 slovenština .....	58
4.5 MORFONOLOGIE .....	58
4.5.1 Vývoj a obsah morfonologie .....	58
4.5.1.1 definice .....	58
4.5.1.2 morfoném .....	58
4.5.1.3 morfofoném .....	58
4.5.1 Hláskové a morfologické alternace .....	58
4.5.1.1 hláskové alternace .....	58
4.5.1.2 morfonologické alternace / střídání fonémů .....	59
4.5.1.3 alternační řada .....	59
4.5.1.3.1 uspořádanost × neuspořádanost .....	59
4.5.1.3.2 základní varianta .....	59
4.5.1.3.3 princip většinové podoby .....	59
4.5.1.3.4 fonémová nula / nulový morf. ....	59
4.5.1.3.5 alternativy korelativní, disjunktní .....	59
4.5.1.3.6 typy alternací .....	59
4.5.1.4 nerovnocennost alomorfů .....	59

4.6 SHRNUTÍ.....	59
<b>5 DOSAVADNÍ EMPIRICKÁ (ZEJM. LEXIKOGRAFICKÁ) ZPRACOVÁNÍ MORFEMATIKY ...61</b>	
5.1 NÁSTIN VÝVOJOVÉHO DRIFTU .....	61
5.2 PŘEHLED VYBRANÝCH LEXIKOGRAFICKÝCH PŘÍSTUPŮ K MORFEMATICE.....	62
5.2.1 <i>Čeština</i> .....	62
5.2.1.1 Slavičková: Retrogradní morfematický slovník češtiny (1975) .....	62
5.2.1.2 Šiška: Bázový morfematický slovník (2005) .....	65
5.2.2 <i>Slovenština</i> .....	67
5.2.2.1 Morfematický slovník slovenčiny (1999) .....	67
5.2.2.2 Slovník koreňových morfém slovenčiny (2005) .....	68
5.2.3 <i>Polština</i> .....	69
5.2.3.1 Słownik gniazd słowotwórczych współczesnego języka ogólnopolskiego (2001–2004).....	69
5.2.4 <i>Ruština</i> .....	70
5.2.4.1 Russian Derivational Dictionary (1970) .....	70
5.2.4.2 Slovar' morfem russkogo jazyka (1986).....	71
5.2.4.3 Tichonov: Morfemno-ortografičeskij slovar' russkogo jazyka (1996).....	72
5.2.4.4 Chronological Morphemic and Word-Formational Dictionary of Russian .....	73
5.2.5 <i>Ukrajínština</i> .....	73
5.2.5.1 Poljuha: Morfemnyj slovar' (1983) .....	73
5.2.5.2 Karpilovs'ka: Korenevyj gnizdovyj slovnyk ukrajinskoji movi (2002).....	74
5.2.6 <i>Němčina</i> .....	75
5.2.6.1 Augst: Lexikon zur Wortbildung: Morpheminventar (1975) .....	75
5.3 SHRNUTÍ.....	76
<b>6 METODA A METODIKA KOMPLEXNÍ MORF(EMAT)ICKÉ ANALÝZY .....78</b>	
6.0 .....	78
6.1 METODY RUČNÍ MORFEMATICKÉ ANALÝZY .....	79
6.2 AUTOMATICKÉ (A KOMBINOVANÉ) METODY MORFEMATICKÉ ANALÝZY .....	82
6.2.0 <i>Motto</i> .....	82
6.2.1 <i>Přehled počítačových přístupů k morfematické segmentaci</i> .....	82
6.3 <i>Univerzální metoda počítačové reprezentace morfematické analýzy</i> .....	85
6.4 SHRNUTÍ.....	86
<b>7 MORFEMATICKÁ GRAMATIKA A MORFEMATICKY ORIENTOVANÁ SLOVOTVORBA ...87</b>	
7.1 STRUČNÝ POPIS POUŽITÝCH DAT A METOD .....	87
7.2 KVANTITATIVNÍ CHARAKTERISTIKY APELATIVNÍHO SYSTÉMU ČEŠTINY VE SVĚTLE KORPUSOVÉ EVIDENCE.89	
7.2.1 <i>Souhrnná charakteristika systému (RMSČ) prizmatem úzu (synČNK)</i> .....	89
7.2.1.1 Podle počtu slov, slabik a morfémů .....	89
7.2.1.2 Podle podílu apelativ a proprií.....	89
7.2.1.2 Podle zastoupení nesložených slov a kompozit .....	89
7.2.1.3 Podle počtu všech homonym .....	90
7.2.1.4 Podle množství homonym s odlišnou segmentací.....	90
7.2.1.5 Podle podílu jednoznačné a nejednoznačné segmentace (perintegrace) .....	90
7.2.1.6 Podle pokrytí korpusů slovní zásobou RMSČ .....	90
7.2.1.6.1 RMSČ a korpus SYN.....	90
7.2.1.6.2 RMSČ a korpus SYN2009PUB .....	90
7.2.1.6.3 RMSČ a korpus synČNK.....	90
7.2.2 <i>Struktura slovní zásoby podle číselného indexu z RMSČ</i> .....	90
7.2.3 MORFEMATICKÁ STRUKTURA ČEŠTINY PODLE POČTU MORFÉMŮ.....	91
7.2.3.1 Celkový přehled.....	91
7.2.3.2 Podle počtu morfémů a slovních druhů .....	92
7.2.4 <i>Morfematická struktura češtiny podle počtu slabik</i> .....	93
7.2.4.1 Celkový přehled.....	93
7.2.4.2 Podle počtu slabik a slovních druhů .....	93

7.2.5 MORFEMATICKÁ STRUKTURA ČEŠTINY PODLE SLOVNÍCH DRUHŮ .....	94
7.2.6 <i>Morfematická struktura češtiny z hlediska typologie morfémů</i> .....	94
7.2.6.1 Podle pozice ve slově.....	94
7.2.6.2 Podle pozice ve slově a slovního druhu .....	95
7.2.6.3 Podle celé značky s vynecháním slovního druhu .....	95
7.2.6.4 Podle celé značky .....	96
7.2.6.5 50 nejčastějších morfémů podle formy – grafematicky (homonymně, po odstranění značky) .....	97
7.2.6.6 50 nejčastějších morfémů podle formy a pozice ve slově – grafematicky.....	98
7.2.6.7 50 nejčastějších morfémů podle formy a celé značky – grafematicky .....	100
7.2.6.8 50 nejčastějších morfémů podle formy – fonematicky (homonymně, po odstranění značky) .....	101
7.2.6.9 50 nejčastějších morfémů podle formy a pozice ve slově – fonematicky .....	102
7.2.6.10 50 nejčastějších morfémů podle formy a celé značky – fonematicky .....	103
7.2.7 <i>Morfematická struktura češtiny z hlediska fonematického složení (CV)</i> .....	104
7.2.7.1 50 nejčastějších morfémů podle složení z konsonantů a vokálů – grafematicky (forma a celá značka) .....	105
7.2.7.2 50 nejčastějších morfémů podle složení z konsonantů a vokálů – grafematicky (forma) .....	106
7.2.7.3 Morfém z hlediska počtu slabik.....	107
7.2.7.4 50 nejčastějších morfémů podle složení z konsonantů a vokálů – fonematicky (forma a celá značka, C – kons., S – slabikotv. kons., V – jednoduch. vok., D – diftong).....	107
7.2.7.5 50 nejčastějších morfémů podle složení z konsonantů a vokálů – fonematicky (forma a celá značka, C – kons., V – vok.) .....	109
7.2.7.6 50 nejčastějších morfémů podle složení z konsonantů a vokálů – fonematicky (forma, C – kons., V – vok.) .....	110
7.2.8 <i>Morfematická struktura češtiny ze slootovorného hlediska</i> .....	111
7.2.8.1 Vzorce struktury slov podle typů morfémů (preradix. část   radix. č.   postradix. č.) .....	111
7.2.8.2 Vzorce struktury slov podle typů morfémů (preradix. část   radix. č.   postradix. č.).....	115
7.2.8.3 50 nejčastějších vzorců struktury slov podle typů morfémů a slovního druhu (preradix. část   radix. č.   postradix. č.).....	120
7.2.8.4 50 nejčastějších vzorců struktury slov podle typů morfémů a slovního druhu (preradix. část   radix. č.   postradix. č.).....	121
7.2.8.5 50 nejčastějších vzorců struktury slov podle celé značky s vynecháním slovního druhu (preradix. část   radix. č.   postradix. č.).....	122
7.2.8.6 50 nejčastějších vzorců struktury slov podle celé značky s vynecháním slovního druhu (preradix. část   radix. č.   postradix. č.).....	123
7.2.8.7 50 nejčastějších vzorců struktury slov podle formy, celé značky a slovního druhu (preradix. část   postradix. č.).....	125
7.2.8.8 50 nejčastějších vzorců struktury slov podle formy, celé značky a slovního druhu (preradix. část   postradix. č.).....	126
7.2.9 <i>Morfematické alternace</i> .....	127
7.3 PREDIKCE KVANTITATIVNÍCH CHARAKTERISTIK PROPRIÁLNÍHO SYSTÉMU ČEŠTINY .....	127
7.4 SHRNUTÍ.....	129
<b>8 NÁVRH EFEKTIVNÍ MORFEMATICKÉ SYNTÉZY .....</b>	<b>130</b>
8.0 MOTTO .....	130
8.1 OBECNÁ VÝCHODISKA.....	130
8.2 ARCHITEKTURA NAVRHOVANÉHO SYSTÉMU.....	132
8.3 SHRNUTÍ.....	136
<b>9 VÝZNAM JAZYKOVÉ UNIVERZÁLIE V GLOBALIZOVANÉM SVĚTĚ (ZÁVĚREM).....</b>	<b>137</b>
<b>RESUMÉ .....</b>	<b>139</b>
<b>SEZNAM KLÍČOVÝCH SLOV .....</b>	<b>139</b>
<b>LITERATURA.....</b>	<b>140</b>
<b>SEZNAM OBRÁZKŮ A TABULEK.....</b>	<b>159</b>
<b>PŘÍLOHA</b>	
1.0 PŮVODNÍ ZNĚNÍ NÁVRHU PROJEKTU <i>MORFEMATICKÁ ANALÝZA ČEŠTINY</i>	



1.1 POSUDEK Č. 1 [ZTAJENÝM AUTOREM ZŘEJMĚ J. ŠIMANDL]

1.2 POSUDEK Č. 2 [NEODTAJNĚNÝM AUTOREM NEPOCHYBNĚ K. PALA]

„Dnes bychom mohli říct, že se některá témata jeví spíše jako vyčerpaná než jako neplodná, a proto se jim málokdo věnuje (morfematika a snad i fonologie). Teď na ně sice doba není, ale to neznamená, že v budoucnu nedojde k nějakému impulsu, který způsobí, že se znovu dostanou do centra zájmu.“<sup>+</sup>

(Iva NEBESKÁ)

„V případech, kdy při identifikaci morfémů dochází ke kolísání, kdy se rozhodujeme mezi několika možnostmi členění, se projevuje určitá vágnost jazykového systému (...).“<sup>++</sup>

(Eleonora SLAVÍČKOVÁ)

„Člověk nežije v holé empirii, není se svým životem uzavřen do mezí ověřených, smyslových dat. Má také ještě úzkost, touhu, stesk, zvědavost, fantazii, potřebnost a sen – a těmi co chvíli přeskakuje věcné údaje smyslů, logiku rozumu i sebe samého.“<sup>+++</sup>

(Emanuel FRYNTA)

„Ale ten výborný jazyk tak hojný jest.“

(Josef JUNGSMANN)

---

<sup>+</sup> Lehečková, E. – Chromý, J. *Rozhovory s českými lingvisty*. Praha: Dauphin, 2007, s. 193. ISBN 978-80-7272-8.

<sup>++</sup> Slavičková, E. *Retrogradní morfematický slovník češtiny*. Praha: Academia, 1975, s. 26.

<sup>+++</sup> Frynta, E. *Zastřená tvář poezie*. Praha: Nakladatelství Franze Kafky, 1993. Kap. 10, Smysl nesmyslu (Nonsens), s. 129. ISBN 80-900609-6-X.

## 0 Exolingvální prolegomena

### 0.1 Jazyková infrastruktura jako pojítko mezi tradicí a inovací

Nejen v lingvistice patří k nejoceňovanějším vlastnostem každého badatele akribie a mezi nejdůležitější kritéria posuzování výsledků jeho práce kredibilita. Jak ale uvádí, poněkud vyhrcočně, Walter J. Ong: „*Neexistuje způsob, jak přímo vyvrátit to, co text tvrdí. I poté, co jej naprosto a drtivým způsobem zpochybníme, říká úplně totéž, co předtím.*“ (ONG 2006: 94) Jinými slovy, veřejné knihovny jsou dnes mnohdy plné titulů s nevalnou pověstí, o které však téměř nikdo nestojí, stejně tak existují hodnotná díla, jež byla často z ideologických, ale pohříchu i osobních důvodů vymazána z pomyslné mapy lidského poznání. Odstranění jmen a jejich prací z psaných i nepsaných dějin jednotlivých oborů je o to snazší, nebyl-li jejich význam potvrzen tiskem. Díla, která přežívají – často v neúplných autografech a opisech – uložená v hloubi archivů a depozitářů, není těžké podle momentálně vůle a potřeby oslavit nebo zatratit publikovaným svědectvím.

Tam, kde existují pouhé domněnky a nepřímé důkazy (např. Komenského *Thesaurus*), nebo v případě vžitých hodnocení (barokní „temno“) či pohodlně reprodukováných názorů (pokrokový mučedník Bolzano a plachý šovinista Jungmann), vzniká mnohdy scestný dojem redundance primárních pramenů. Tím spíše, že značná část kanonických textů není psána v obecně srozumitelném, vernakulárním jazyku, ale v starší bohemistice převážně latinsky a německy<sup>1</sup>, dnes častěji anglicky. Novější se pak stává synonymem pro lepší a dokonalejší, texty vydané před jakoukoli zásadnější pravopisnou reformou získávají v očích laické veřejnosti automaticky (před)historický ráz, a spolu s formou je tak antikvován i obsah. Podobně absurdní je i hledání viníků čehokoli, jejichž označení v té či oné věci se potom stává jejich univerzálním atributem petrifikovaným zejména školní výukou.

Václav Hanka je tak zodpovědný za národní rukopisnou ostudu, ač jako dílčí pravopisný reformista a editor staročeských památek by byl pravděpodobně chválen. Václav Jan Rosa zase nechvalně proslul jako komický neoterista, který však naštěstí češtinu příliš nepoznamenal, zatímco jeho průkopnické poznatky o českém vidovém systému a inovace v oblasti lexikografie se do širšího povědomí bohužel nedostaly. Ani Josefa Dobrovského, kterému v myslích většiny Čechů slouží ke cti snad jen rozpoznání rukopisných falz, neuchránila před odsudky – překvapivě i ze strany některých jazykovědců – např. pronikavá analýza slootovorného systému nebo snaha modernizovat pravopisnou soustavu. Potíže s tzv. českou diglosií (srov. např. BERMEL 2010a) jsou zkrátka připisovány výlučně jemu.

Z hlediska hlavního tématu této práce, tedy morfematické problematiky, je nanejvýš smutné, že proskribování některých autorů i jejich děl přetrvalo politické změny po roce 1989, což postavilo nečetné konciliantní snahy o velkorysost příslušející „vítězům“ (srov. LEŠKA 1993) do značně nepříznivého světla. Jak si např. povšiml A. Stich (2001), z „*této základní memoárové stati* [TĚŠITELOVÁ 1999, pozn. J. L.] *zmizelo kardinální dílo oboru (zpracovávaného za vedoucí účasti hlavního oborového normalizátora), a co ještě hůř, že paralelně (!) [zvýraznil A. S.] odsud zmizela jména dvou žen, jejichž dílo bylo po roce 1970 rozmetáno, znevaženo a uvedeno v nic. První jméno zní Eleonora Slavičková (byla i bibliografkou oboru kvantitativní lingvistika), ta druhá se jmenovala JITKA ŠTINDLOVÁ (...).*“ (ibid.: 221n)“ Stichův příspěvek byl motivován přáním připomenout význam druhé jmenované, předkládaná práce je mutatis mutandis zase obdivným navázáním na výsledky morfematického výzkumu **Eleonory Slavičkové** (\* 18. 9. 1936), který byl navzdory násilnému přerušení koncem šedesátých let – pro svou zjevnou nenahraditelnost<sup>2</sup> – v roce

<sup>1</sup> Jazykovědec a literární historik Alexandr Stich (1934–2003) přirovnával s patřičnou dávkou nadsázky chybějící znalosti latiny a němčiny u bohemistů-profesionálů k absenci jedné či obou dolních končetin.

<sup>2</sup> Okolnosti vydání nejsou známy, přesto není možné vyloučit, že za překvapivým vydáním slovníku několik let po autorčině vyloučení z oboru stojí drobná poznámka z úvodu k dílu, kde se píše: „*Zatím však byl zpracován –*

1975 završen vydáním *Retrográdního morfematického slovníku češtiny* (viz 5.2.1.1). Náklad díla čítal na tehdejší poměry nepatrných 700 výtisků.

Najít společné téma, které by přesvědčivě charakterizovalo dění v jazykovědné bohemistice posledních dvou decenníí, není právě snadné. Obnovovala, spojovala a proměňovala se tradiční univerzitní i akademická pracoviště, vznikala nová (např. *Ústav Českého národního korpusu*, dále jen ÚČNK), navazovala se přerušená činnost spolková (např. obnovený *Pražský lingvistický kroužek*) i výzkumná (např. *Ústav formální a aplikované lingvistiky* na MFF UK nebo fonetický výzkum na FF UK). Pořádání tematicky rozmanitých národních i mezinárodních konferencí se stalo trvalou součástí odborného lingvistického života. Největší změna se ovšem odehrála v oblasti publikační a komunikační. Prudký nárůst výpočetních, zobrazovacích i kapacitních možností počítačové techniky spolu s postupným propojováním pracovišť i jednotlivců prostřednictvím internetu předznamenal **budování jazykové infrastruktury**, tak charakteristické pro počátek 21. století.

Převážně díky veřejné podpoře začaly vznikat digitální knihovny<sup>3</sup>, jazykové korpusy<sup>4</sup> (viz např. ČERMÁK 1995 a 2008b) nebo bibliografické databáze<sup>5</sup>. Postupně jsou zpřístupňovány rovněž elektronické verze historických i novodobých slovníků<sup>6</sup> a také excerpčních kartoték. Moderní technika tak velmi efektivně zprostředkovává bez časových a geografických omezení mj. primární prameny, které byly ještě donedávna nedostupné. Toto úsilí se pochopitelně neobejde bez často zbytečných chyb a omylů, podstatná je však otázka, jestli dosavadní zvyky (ostatně předávané a upevňované celými vědeckými generacemi učitelů a žáků), totiž vycházet většinou ze zprostředkovaných informací a soudů, bude alespoň odborná veřejnost ochotná změnit. Snad je to záležitost delšího časového vývoje, možná schází dosud decentralizovanému vytváření infrastruktury vhodný způsob popularizace (srov. např. LEBEDA 2009). Mé dosavadní zkušenosti s provozem elektronických verzí českých morfematických slovníků<sup>7</sup> jsou však bohužel důvodem ke skepsi<sup>8</sup>, zvláště když se i v nejnovějších pracích objevují nepřesnosti<sup>9</sup> (např. ČERMÁK 2010), které by snadno verifikovalo lepší využití dnešních možností.

Apel na zajištění lepší informovanosti o tom, co je dnes k dispozici a jaké údaje se kde (a s jakou kredibilitou) dají získat (srov. zejm. ŠIMANDL 2010), má o to větší důležitost, že se některé datové zdroje stávají ne neproblematickým východiskem tištěných publikací (nejnověji CVRČEK ET AL. 2010) a že se i pro češtinu začínají demokratizovat tzv. webové korpusy (srov. SPOUSTOVÁ – SPOUSTA – PECINA 2010 nebo POMIKÁLEK – RYCHLÝ – KILGARRIFF 2009). Zdá se, jako by právě internet a v posledních letech jistým způsobem také televizní, tzv. veřejnoprávní zpravodajství (srov. např. ADAM 2007) přispívaly k pocitu, že bude Josef Dobrovský po dvou stoletích konečně vyviněn, změní-li se dichotomie *spisovný* a

---

*pokud je nám známo – pouze jediný morfematicky členěný, ne však retrográdně uspořádaný slovník, a to pro ruštinu („Russian Derivational Dictionary“), který zpracoval na kalifornské univerzitě v Los Angeles /USA/ prof. D. S. Worth s kolektivem.*“ ([RMSČ] SLAVÍČKOVÁ 1975: 12) [zvýraznil J. L.] Nebyl by to jistě první ani poslední projev „studené války“ v československé vědě.

<sup>3</sup> Nejvýznamnější je v tomto ohledu činnost *Národní knihovny ČR*, viz např. <<http://www.manuscriptorium.eu/>> nebo <<http://kramerius.nkp.cz/>>. V mezinárodním měřítku je i pro české prameny nezastupitelné digitalizační úsilí vyvíjené společností Google: <<http://books.google.cz/>>.

<sup>4</sup> Zejm. <<http://www.korpus.cz/>> a <<http://vokabular.ujc.cas.cz/banka.aspx>>.

<sup>5</sup> Např. <<http://www.ucl.cas.cz/cs/databaze/>>.

<sup>6</sup> Za všechny např. <<http://vokabular.ujc.cas.cz/zdroje.aspx>> nebo <<http://www.slownjk.cz/>>.

<sup>7</sup> <<http://www.morfemy.cz/>>.

<sup>8</sup> Za více než rok a půl existence digitálních edic se objevil jen jediný zájemce (přístup ke slovníkům je podmíněn e-mailovou žádostí, sumární statistický přehled je volně dostupný).

<sup>9</sup> Údaj o 1737 kořenech (tj. neuváděných 3894 alomorfů a variant), který F. Čermák (např. 2010: 197) přebírá ze Slavičkové (RMSČ 1975), se týká pouze stylisticky neutrální slovní zásoby a nikoli celého (domácího) lexika. Celkový počet kořenových alomorfů a variant je v širším jádru slovní zásoby (tj. mezi cca 63 tisíci lemmaty) zhruba 11 tisíc.

nespisovný v distinkci *dominantní* (většinový) a *minoritní* (viz CVRČEK – VYBÍRAL 2005 a především CVRČEK 2008 a CVRČEK ET AL. 2010). Představa češtiny ocitající se vně axiologické soustavy naráží nejen na rezultující neformalizovatelnost (a také nepochopitelnost) vysoce variantní a takto redundantní české flexe, tedy např. v protikladu k algoritmizaci představené v osmé kapitole této práce, ale katexochén na podobně organizovanou soustavu pravidel společenské etikety, o níž je v současné době mezi lidmi velký zájem (viz např. MATHÉ – ŠPAČEK 2010).

## 0.2 Bádání nezatížené staršími přístupy k jazyku a tzv. bezhodnotová lingvistika

Z hlediska publikační činnosti patří tzv. korpusová lingvistika v jazykovědné bohemistice mezi nejrychleji se rozvíjející odvětví. Problém však spočívá mj. v tom, že jsou mnohdy za korpusové pokládány jakékoli práce, při jejichž přípravě badatel využíval nějaký počítačový korpus přirozeného jazyka. Jde vlastně o nepříznávanou podvojnost, kdy je zatemňována hranice mezi přístupem k jazykovým datům označovaným jako *corpus-based* (tzn. ověřování hypotéz na korpusových datech) a *corpus-driven* (tj. analyzování obsahu korpusů bez apriorních soudů) (porovn. ŠIMANDL 2010: 230–232). K prvnímu pracovnímu postupu se hlásí mnoho lingvistů, těch, již vyznávají druhý přístup, je v českém prostředí nevelké množství. Přesto jsou to právě oni, kteří se vytrvale snaží přesvědčit ostatní (včetně laické veřejnosti), že se jejich způsob práce stal samostatnou jazykovou disciplínou, navíc takovou, která překonává vše, co tu bylo před ní.

Na zadní straně obálky *Studií z korpusové lingvistiky*, které od roku 2006 vydává ÚČNK, tak např. stojí, že „edice (...) předkládá (...) **nejnovější a staršími přístupy nezatížené** výsledky bádání v relativně mladém **oboru** lingvistiky.<sup>10</sup>“ [zvýraznil J. L.] Podobně běžný je obrat „korpusová a tudíž objektivnější data“ (ČERMÁK 2010: 66). Ani odpůrci či přesvědčení skeptikové ovšem nejsou mezi uživateli korpusů se svými názory osamoceni. L. Veselý např. argumentuje takto: „Když už chce být korpusová lingvistika lingvistickou disciplínou – a já myslím, že jí není (jinak bychom měli i lingvistiku manuálně excerpční nebo lingvistiku dotazníkovou) –, neměla by se omezovat jen na to, že bude počítat, kolik je čeho v korpusu a kolikrát toho je víc nebo míň než něčeho jiného. To by pak místo studií o jazyce produkovala jen jakési registry jazykových jednotek s informacemi o četnosti jejich výskytů. Bylo by to příliš pracné a bylo by to příliš málo.“ (VESELÝ 2008: 216) Bez ohledu na to, co stojí na obálce edice, píše v podobné duchu i Šimandl: „Pro poznání jazyka a způsobů, jakým se s ním zachází, bude plodnější nepřekrýt tuto různost sumární statistikou, ale mapovat ji a hodnotit se stálým zřetelem k této různosti (...).“ (ŠIMANDL 2010: 330)

Také pro předkládanou práci jsou údaje obsažené v korpusech důležité, i když zkoumání morfematické stavby češtiny se od ostatních analýz korpusového materiálu v mnoha ohledech liší. Pokud je např. pro morfologii nezbytné přihlížet k jazykovým kvalitám jednotlivých textů (ibid.: 229), morfematická dekompozice slovních tvarů se nemění, ať jsou kvantitativní nebo komunikačně-pragmatické hodnoty výchozích dat jakékoli. Hranice mezi morfy (viz kap. 6) a jejich (hierarchické) členění (viz kap. 4) je při identitě tvaru stejné v psaném i mluveném (transkribovaném) textu, při spisovném i nespisovném (příznakovém i bezpříznakovém) užití jazyka atp. Z čistě epistemického hlediska si morf(emat)ický výzkum vystačí s pouhými frekvenčními seznamy, je tak eo ipso korpusový, tedy ve smyslu metody *corpus-driven*. To se ovšem týká jen výchozího materiálu, samotnou analýzu (viz 6.2) už – navzdory četným pokusům<sup>11</sup> – korpus sám řídit nedokáže (srov. např. STATČ 2009: 123–128).

Více než spor o to, jaký je vlastně (sociální) status korpusové lingvistiky, je důležité nepřehlédnout opakovaně podsouvané tvrzení o nezatíženosti korpusových výzkumů

<sup>10</sup> Citováno podle 10. svazku řady, ŠIMANDL 2010.

<sup>11</sup> Pro češtinu zejm. CVRČEK 2009.

výsledky starších přístupů. Pokud by to měla být pravda, dala by se např. z libovolného korpusu, tedy užití jazyka (*parole*) vyvodit stejná abstrakce systému (*langue*), což je nejen vzhledem k rozdílům ve velikostech korpusů<sup>12</sup> zcela nemyslné. Ilustrovat to může např. údaj o počtu substantiv v 500milionovém korpusu SYN (viz pozn. 12), která mají kompletně doložená paradigmata (tzn. všech 7 pádů krát obě dvě čísla). Ze všech obsažených substantivních lemmat (649 571) je to pouhých 413, tedy jen 0,064 %. Ještě přesvědčivějším dokladem toho, že teze o nezátíženosti je myšlena pouze proklamativně, je např. opakované citování údajů týkajících se systému jazyka, reprezentovaného i v roce 2010 slovníkem SSJČ (<sup>2</sup>1989, <sup>1</sup>1960–1971), které uvádí Čermák (2010: 156, 163, 234). Zvláště pozoruhodné je jméno autorky přebíraných kvantitativních údajů, jímž je Jitka Štindlová (1928–1989), průkopnice zpracování češtiny mechanograficky, tj. pomocí tzv. děrných štítků (srov. např. ŠTINDLOVÁ 1961 a 1968).

Neméně významným svědectvím o závislosti české korpusové lingvistiky na starších výsledcích je automatická *morfologická analýza* (přiřazení možných gramatických a lexikálních interpretací jednotlivým výskytům slovních tvarů v textu) a *lemmatizace* (asociace morfologické formy a reprezentativního tvaru, pro substantiva zpravidla Nsg a pro verba infinitivu), kterou zajišťují programové nástroje postavené na teoretických pilířích předkorpusové gramatiky a lexikografie. Jakékoliv vyhledávání v korpusu je tak poznamenáno formalismem, jenž např. rozhoduje, jaké slovní tvary a varianty jsou sjednoceny pod společné lemma<sup>13</sup>. V tomto směru se příliš neliší oba nejužívanější systémy, a sice tzv. pražský<sup>14</sup> – využívaný pro korpusy ÚČNK – a tzv. brněnský<sup>15</sup> (viz kap. 8.2). Veškeré korpusové publikace (např. STATČ 2009) tak hodnotí jazykovou realitu prizmatem „historického“ *morfologického značkování* (výsledek morfologické analýzy zjednotněný na základě statistického a/nebo pravidlového přístupu, tj. výběr jedné z plauzibilních interpretací).

V průběhu posledních let přestávají být korpusy češtiny – řečeno s F. X. Šaldou – pouhým patosem a inspirací, ale stávají se namnoze jediným kritériem, které má – zatím jen v očích zastánců tohoto přístupu – sloužit při volbě jazykových prostředků, a to jako alternativa k dosavadním preskriptivním příručkám (viz hlavně CVRČEK 2008). Jde tu především o oblast ortografie a formální morfologie. Jazyková zjištění nejsou nijak interpretována, tedy opatřena příznakem přijatelnosti pro různé komunikační situace, namísto toho je buď podán celý inventář prostředků opatřený statistikou výskytu (např. STATČ 2009) nebo je provedeno srovnání frekvenční distribuce vybraných výrazových konkurentů (CVRČEK ET AL. 2010). V reakcích široké veřejnosti se pro tuto lingvistickou adaptaci principu „volného trhu“ začíná vžívat termín *bezhodnotová lingvistika*<sup>16</sup>.

Přítomná práce se v tomto ohledu vědomě staví na stranu lingvistické tradice, k čemuž ji nutí např. – pro zodpovědnou morfematickou analýzu imanentní – závislost na etymologických výsledcích (viz 6.1). Stejně tak se hrdě hlásí k poznatkům z oblasti formální morfologie a k výsledkům mnohasetleté lexikografické činnosti, bez nichž by modelování systému (viz 7.2 a 8.2) (zejm. ve sféře potenciality) nebylo vůbec možné. Axiologický rozměr užívání jazyka stojí mimo operační pole, do kterého spadá morfematická analýza. Její přístup

<sup>12</sup> Např. (rozsahem) milionový korpus mluvené češtiny (z Čech) ORAL2008 vs. 700milionový korpus psané publicistiky SYN2009PUB. Podrobněji o složení viz: <<http://ucnk.ff.cuni.cz/struktura.php>>

<sup>13</sup> U adjektiv a adverbii např. komparativy a superlativy pod lemmatem odpovídajícím pozitivu, jmenňé adjektivní tvary pod složenými (*hotov* pod *hotový*), n-ová participia pod verbálním lemmatem (*koupen* pod *koupit*) atd. Mnoho problémů se týká určování gramatických kategorií sloves (zejm. vidu a času) a obecně analytických tvarů, příp. viceslovných lexikálních jednotek (dosud prakticky neřešený nedostatek).

<sup>14</sup> Viz HAJÍČ 2004.

<sup>15</sup> Viz např. SEDLÁČEK 2004.

<sup>16</sup> Doklady lze nejpohodlněji najít v čtenářských diskuzích pod příspěvky jazykového blogu V. Cvrčka, viz <<http://blog.aktualne.centrum.cz/blogy/vaclav-cvrcek.php>>.

je do jisté míry synkretický, protože – v pojetí této práce – využívá všechny dostupné klasifikační systémy a zároveň není na žádném z nich závislá (viz 8.2).

### 0.3 Homo theoreticus a homo practicus

Diapazon lingvistických teorií, a to jak obecných, tak i vysoce specializovaných, příkře kontrastuje se skutečností, že prakticky neexistují konsenzuální definice těch nejjobecnějších jazykovědných pojmů, jakými jsou např. *slovo*, *věta* nebo *text*. Příčinou může být buď indexální, deiktická povaha označovaných jevů (např. individuálně a dobově proměnná hranice mezi *synchronií* a *diachronií*<sup>17</sup>), častěji jde však o pouhý problém výrazové delimitace obsahového kontinua, o jehož ontologickém statusu není pochyb (např. poloha slabičného předělu vs. schopnost určit počet slabik). Nedefinovatelnost konkrétních pojmů se ovšem může stát přímým podnětem pro vznik teorie, která axiomaticky popře existenci substance. V české lingvistice je to případ Cvrčkova *Konceptu minimální intervence* (CVRČEK 2008), který staví na vágnosti pojmu *spisovnost*.

Důvodem nezdaru nebo krátkého života některých teoretických konceptů<sup>18</sup> může být i někdy i to, že „[p]ožadavek vyčerpávajícího popisu je nadřazen požadavku jednoduchosti“ (HJELMSLEV 1972: 15). Se vzrůstající složitostí teorie (ať už definitorickou nebo konceptuální) se zpravidla snižuje její aplikovatelnost. U *Funkčního generativního popisu* (viz např. HAJIČOVÁ – PANEVOVÁ – SGALL 2002), který je dnes značně rozsáhlou metateorií vícerovinné závislostní analýzy jazyka (češtiny), se např. uplatnění zásad neobejde bez obsažných manuálů a dlouhodobého výcviku anotátorů. I tak se jakýkoli nejasný případ – v autentických jazykových datech – může stát (i rekurentně) popudem pro doplnění manuálů i úpravy teoretických východisek.

Volba tématu, kterým je v mém případě *morfematická struktura češtiny* (tedy systémová i textová identifikace a analýza morfémů<sup>19</sup>), nutně otevírá otázku jeho vhodného teoretického uchopení. Za více než sto let, po kterou termín *morfém* existuje (viz kap. 3), se podařilo promítnout jeho konceptuální jádro (tedy zpravidla korelaci mezi částmi slova a jejich invariantním lexikálním a/nebo gramatickým významem) do teoretického rámce většiny tradičních lingvistických disciplín. Dlouhou dobu byl morfém centrální jednotkou morfologie (srov. např. BEDNAŘÍKOVÁ 2010) a hyperonymickou jednotkou (mor)fonologie, dnes nachází své místo – prakticky rovnocenně – ve fonologii, morfologii, slootovorbě (někdy označované jako derivační morfologie), v syntaxi i lexikologii. S postupnou změnou morfologického paradigmatu, kdy začal morfém v pozici ústřední jednotky ustupovat slovu (tvaru) nebo lexému (ibid.), se jako samostatná disciplína – podle A. Gerda (GERD 2004) zhruba od osmdesátých let – počíná prosazovat *morfematika*, resp. *morfotaktika*. Název je odvozen od hlavní jednotky, proto je její pozice zjevně neotřesitelná.

Kromě pochopitelného zaměření morfematiky na zkoumání morfému v paradigmatickém (delimitace a funkční rozlišení jednotek) i syntagmatickém (kombinatorika morfémů, terminologicky často jako *morfotaktika*), se předmět nedlouho osamostatněného oboru profiluje značně interdisciplinárně. Pokud pomineme Horeckého *Morfematickou strukturu slovenčiny* (HORECKÝ 1964), která je ještě zaměřena spíše slovotvorně a formálněmorfologicky (ve smyslu flektivní paradigmaticky), neexistuje zatím žádná plnohodnotná monografie. Bosákova rozsáhlejší studie *Východiská morfémové analýzy* (BOSÁK – BUZÁSSYOVÁ 1985) má už podle názvu zřetelné praktické zaměření, jakkoli

<sup>17</sup> Viz VEČERKA 2004.

<sup>18</sup> František Daneš se k této problematice vyjádřil v neformálním rozhovoru takto: „*Je ale opravdu podivuhodné, kolikrát se Chomsky za dob svého působení v lingvistice proměnil. Paul Garvin mi tohle velmi dobře vysvětlil: „Viš, Franto,“ říkal mi, „v Americe musíš každých pět let přijít s něčím novým, jinak jsi mrtvý muž. Podíváte-li se na Chomského, on tuhle podmínku naprosto splňuje.“*“ (CHROMÝ – LEHEČKOVÁ 2007: 47n)

<sup>19</sup> K definici pojmu viz kap. 4, o metodách delimitace viz kap. 6.

obsahuje i cenné teoretické poznatky a pečlivou analýzu starších pramenů. Výchozí jazykový materiál se navíc omezuje jen na tři, třebaže dominantní slovní druhy: substantiva, adjektiva a slovesa, kromě toho autor abstrahuje od segmentace tzv. mezinárodních slov. Jak GERD (2004), tak i HRAJZNUCHINA ET AL. (1979) jsou jen souborná vydání heterogenních studií, v nichž se kromě tradičních témat častěji uplatňují také algebraické, formálnělogické anebo statistické přístupy. Bergenholtz se ve svém díle *Zur Morphologie deutscher Substantive, Verben und Adjektive* (BERGENHOLTZ 1976) věnuje – v teoretickém rámci „komunikační gramatiky“ – prakticky jen empirické analýze textových pramenů z hlediska morfému a jeho slovnědruhové (ne)vyhraněnosti. Ostatní práce většího rozsahu se morfematickou problematikou zabývají buď jen v kontextu svébytných lingvistických teorií a směrů, případně volí postup *pars pro toto*.

Simenon Romportl (1970) zkoumá na gramatické složce syntetických slovesných tvarů interakci mezi výrazem (Saussurovo *signifiant*) a významem (*signifié*), Kořenský (1972) podává morfosyntaktický popis „realizace pádových vzorců [substantiva] ve výpovědních schématech češtiny“, v němž akcentuje výrazovou stránku nominální flexe. V. Straková se – ve srovnání ruštiny a češtiny – zaměřila na komplexní zhodnocení substantivní derivace (STRAKOVÁ 1973) a věnovala se také sémantické stránce morfémů, konkrétně experimentálnímu nástínu derivační typologie (STRAKOVÁ 1985). Minuciózní analýzu kmenotvorných přípon (ruského) slovesa s cílem potvrdit jeho morfémový status uskutečnila Z. Skoumalová (1976). Marvanova monografie o českém stupňování (MARVAN 2008), které pokládá za součást flexe, nikoli slovtvorby, podává tento jev v širokých slavistických a dokonce v (indo)evropských souvislostech. S návrhem sloučení metodiky morfematické i slovtvorné analýzy přichází Bednaříková (2009), sémantickou složku kompozit analyzují z hlediska kombinací domácích a cizích základů práce P. Mittera (MITTER 2003b, 2006 a 2008).

Pozici morfému v novějších morfologických směrech objasňuje monograficky např. Beard (1995) nebo Aronoff (1994), z hlediska generativních teorií je úloha morfému podrobně zkoumána ve slovtvorbě (ARONOFF 1979), morfosyntaxi (o tzv. distribuované morfologii česky např. ZIKOVÁ 2007) anebo ve fonologii (ZIKOVÁ 2008). Své místo má morfém také v komplexních teoriích, jako je *Funkční generativní popis* (zde však jen proklamativně) nebo sémanticky orientovaná *Mean-Text theory* (stručný popis např. v MEL'ČUK 2006). O morfému z hlediska lexikální sémantiky píše např. Čermák (2010). Nejvýznamnější teoretické práce o morfematické stavbě češtiny publikoval Miroslav Komárek, jehož rozšířené vydání *Příspěvků k české morfologii* (KOMÁREK 2006) jednak představuje některé podnětné myšlenky v nezkrácené podobě – tedy např. ve srovnání s druhým dílem *Mluvnice češtiny*<sup>20</sup> (MČ2) –, především však plní funkci neformální morfematické monografie. Zásluhu na tom mají kromě základní terminologické stati (KOMÁREK 2006c) především cenné studie/kapitoly o hodnocení kmenotvorných přípon u sloves (KOMÁREK 2006b, d, i), stavbě adjektivní a pronominální flexe (KOMÁREK 2006g) a o problematice vidu (KOMÁREK 2006 e, f).

Představu teoretické části modelové monografie dotvářejí časopisecké a sborníkové příspěvky, které se – kromě historiografických exkurzů (SOKOLOVÁ 2008) – věnují např. (systémové) kvantitativní, funkční a sémantické analýze propriálních (OLOŠTIAK 2006) a apelativních kořenů (SOKOLOVÁ – GENČI 2006a, SLAVÍČKOVÁ 1965, SLAVÍČKOVÁ – VAŠÁK 1972), jejich fonematickému složení (HORECKÝ 1969, SOKOLOVÁ – GENČI 2006b, SOKOLOVÁ – IVANOVÁ 2007), vokalickým alternacím (CZÉREOVÁ 2006), alternacím a variantám radixů převzatých slov (ŠIMON 2006) nebo komplexní analýze afixů (IVANOVÁ 2006). Lze se setkat s případovými studiemi o cirkumfixech (ČERMÁK 2008a), redundantních

---

<sup>20</sup> O teoretické nekompatibilitě prvního a druhého dílu viz Bednaříková (2009: 24n).



gramatických morfémech (ŽUROVIČ 1965) nebo o statusu a funkci kmenotvorné přípony ve slovanských jazycích (DOKULIL 1997). Diskutovány byly hranice morfému z hlediska sémantického (SKALIČKA 2004b), typologického (SKALIČKA 2004c) i vzhledem k asymetrii formy a funkce (SKALIČKA 2004a). Ve vztahu k ostatním jednotkám byl prozkoumán vztah morfému a slabiky (SABOL – BÓNOVÁ 2003 a 2004), se zřetelem k ostatním rovinám byly např. rozebírány morfematisco-derivační struktury lexémů (IVANOVÁ – GENČI 2006) i principy kompozice (SOKOLOVÁ – IVANOVÁ – VUŽŇÁKOVÁ 2005, MRAVINACOVÁ 2003, MITTER 2003a, 2004a, b, 2005a, b, 2007).

Kromě teoretických aspektů byla již od počátků věnována pozornost také empirickému výzkumu, a to jak materiálovému (SABOL 1973, 1979), tak především metodologickému (HORECKÝ 1995, SOKOLOVÁ 1994, KRÁLIK 2006), který byl postupně orientován směrem k lexikografickým aplikacím, (SOKOLOVÁ – ŠIMON – MOŠKO 1996, SOKOLOVÁ – IVANOVÁ – DZIAKOVÁ 2003, VUŽŇÁKOVÁ 2006). Při hledání vhodných kritérií segmentace morfému byly kromě jiného diskutovány teorie pravděpodobnosti (distribuce jednotek a jednotlivých přechodů) (SLAVÍČKOVÁ 1962), koncept centra a periférie nebo sociolingvistickým výzkumem zjišťované preference mluvčích (SLAVÍČKOVÁ 1967). Ověřením navrhovaných analytických metod se staly strojově i ručně sestavené morfematiscké a slovotvorné (morfematiscky členěné) slovníky<sup>21</sup>, ať již řazené abecedně (např. SOKOLOVÁ – MOŠKO – ŠIMON – BENKO 1999), retrográdně (např. RMSČ/SLAVÍČKOVÁ 1975), nebo podle kořených morfému (např. WORTH – KOZAK – JOHNSON 1970). Z pochopitelných důvodů (fektivnost) mezi zpracovávány jazyky převažují jazyky slovanské (čeština, slovenština, polština, ruština a ukrajinština), existuje ale také např. slovník pro němčinu (AUGST 1975).

Pokud k popsáným teoretickým směrům, kterými se ubírá současný morfematiscký výzkum, připočítáme rozměr sémiotický (viz kap. 1) a obecnělingvistický<sup>22</sup>, kam spadají i terminologické otázky, otevírá se značně široké pole pro vznik moderně pojaté monografie. Není však jasné, kolik by vynaložené úsilí mohlo – při respektování stávajících parametrů popisu<sup>23</sup> – přinést nového, když už Slavíčková před více než čtyřiceti lety konstatovala, že „[e]xistuje mnoho teoretických definic morfému, ale metodika vymezování konkrétních morfému je specifická problematika, dosud opomíjená“ (SLAVÍČKOVÁ 1967). Odpověď navíc komplikuje o třicet let mladší zkušenost s naplňováním vysloveného deziderata, která říká: „Naším cieľom bolo segmentovať lexikálne jednotky (vrátane prevzatých) na minimálne segmenty tak, aby sme snížili **subjektívnosť segmentácie, vyplývajúcu z nerovnakej jazykovej kompetencie používateľov**, ktorá ovplyvňuje uvedomovanie si súvislosti medzi lexémami, ale **aj subjektívnosť vyplývajúcu z nejednoznačnosti morfémových hraníc** (...).“ (SOKOLOVÁ – MOŠKO – ŠIMON – BENKO 1999: 10) [zvýraznil J. L.]

Podle mého názoru musí výzkum morfematiscké struktury (češtiny) přiznaně postoupit do své třetí fáze (po první teoretické a druhé empiricko-lexikografické), která by mohla nést název *algoritmicko-mimetická*, příp. *algoritmicko-adaptabilní*, přičemž důraz kladu na formalizaci (algoritmus) a mimetičnost (schopnost nápodoby) nebo adaptabilitu (přizpůsobitelnost). Tím, co má být napodobováno, je ontogenetický jazykový vývoj (nabývání, osvojování jazyka), který je po počátečním prudkém rozvoji v různé míře aktivní po celý život jednotlivce. Systém učení však nechápu v jeho konekcionistickém smyslu, tedy jako modelování kognitivních procesů pomocí umělých neuronových sítí (srov. např. PINKER 2008). Naopak, o segmentaci a identifikaci má rozhodovat výlučně *systémovost* (evidence

<sup>21</sup> Podrobnější informace obsahuje kap. 5.

<sup>22</sup> Sem patří kanonické práce jako např. SAUSSURE 1969, BLOOMFIELD 1933/1965, HJELMSLEV 1971, HARRIS 1960, HOCKETT 1967, GLEASON 1969, příp. SOLNCEV 1981 ad.

<sup>23</sup> Tzn. rekapitulace a zpřesňování definic, vymezování se vůči přístupům volícím podle zaměření to či ono řešení z několika možných, závěry vyvozované z vybraných sporných případů, dílčích sond anebo pouhého jádra jazykového systému, příp. jen z lemmat, grafematiscké podoby morfému atp.

v referenčních, zejm. lexikografických popisech a/nebo příslušnost k funkčním třídám a konstrukčním modelům) a zároveň *produktivnost* (frekvenční, resp. pravděpodobnostní ohodnocení výskytu v systému a/nebo v použití jazyka, úzu).

Jakákoli změna (např. nové slovo připouštějící pozměněnou interpretaci stavby slova staršího nebo jeho částí) se pak (sémioticky) promítne do všech meritorních míst celého popisu. To je při nezbytném požadavku konzistence jinými než počítačovými prostředky nedosažitelné. Algoritmickou část této fáze dnes bez potřebného lingvistického zázemí uskutečňují programátoři, mimetický komponent zcela chybí (bývá obcházen tzv. *supervised*, *semi-supervised* nebo *unsupervised* technikami strojového učení<sup>24</sup>). Návrh takového systému, který bude blíže současnému poznání kognitivních procesů (viz kap. 2), využije všech přednostní formálního zpracování a zároveň zůstane lingvisticky adekvátní (viz kap. 6 a 8), pokládám za nejnaléhavější výzvu morfemického výzkumu.

#### 0.4 Synergie, interoperabilita a morfologické konkláve

Oproti teoretickým pracím a konceptualizacím jsou slovníkové aplikace mnohem známější a populárnější (i mezi laickou veřejností<sup>25</sup>). Dostupnost morfemicky analyzovaných dat na internetu, který je čím dál tím častěji využíván i jako velmi efektivní brána k jiným slovníkovým dílům, je však zatím nepatrná<sup>26</sup>. Značné výhody internetu jako média prezentace morfemické struktury jazyka však mají jednu zásadní podmínku úspěšného využití, a sice efektivní organizaci dat. Ta bývá předmětem častých sporů mezi lingvisty a programátory (matematiky), kteří rádi upřednostňují dílčí, dobře a rychle fungující „technická řešení“ před sice teoreticky přiměřenějšími a komplexnějšími, avšak pomalejšími přístupy (srov. HLAVÁČOVÁ 2009a, ŠMERK 2010). Snaha uchopit morfemickou analýzu (a syntézu) formálním způsobem, tak aby mohla být využívána podobně jako analýza morfologická (viz např. HAJÍČ 2004), vyžaduje hned v počátku odpověď na otázku, jestli je schůdnější cesta úpravy stávajícího systému české počítačové morfologie, nebo je-li nutné uvažovat o vytvoření systému nového. Jinými slovy, poskytuje-li stávající algoritmický popis dostatečně robustní vyjadřovací prostředky pro formalizaci na úrovni morfémů.

Obecně platí, že uživatelé jakéhokoli produktu nejvíce ocení jeho snadné ovládání (instruktivnost), dostatečný výkon (účinnost), možnost vícenásobného užití a přizpůsobitelnost individuálním požadavkům. Ten, kdo vznik produktu financuje, zase očekává, že tvůrci budou pracovat *synergicky* (součinně) a že výsledný produkt bude dostatečně *interoperabilní* (např. připravený na různé formáty vstupu a výstupu). Synergii v praxi může ilustrovat příklad zaměstnance přicházejícího s jedním přenosným zařízením odečítat v ohlášený čas spotřebu tepla, vody, plynu i elektřiny. Vývoj v této oblasti může směřovat k automatickému odečtu bez nutné přítomnosti majitele bytu díky radiové technologii. Stav české počítačové morfologie modifikuje uvedený příklad zhruba tak, že odečet provádějí čtyři zaměstnanci vybavení buď blokem, formulářem, přenosným zařízením nebo radiovým terminálem, každý navíc v jiném termínu. Fakticky existují nejméně čtyři<sup>27</sup> komplexní systémy automatické morfologie, dva akademické a dva komerční. Každý využívá trochu jiný formalismus pro analýzu i syntézu, žádné dva neposkytují vzájemně zcela převoditelný formát výstupu, ani jeden neumožňuje nedestruktivní adaptaci systému pro morfemickou analýzu (viz kap. 6 a 8.1). Koexistence tolika řešení nesdílejících vzájemně

<sup>24</sup> Srov. např. BERKA 2003.

<sup>25</sup> Výjimkou je snad jen česká situace, která je charakteristická – až na slovníkovou část *Pravidel českého pravopisu* – velmi rezervovaným vztahem uživatelů jazyka k jednojazyčné lexikografii a jejím produktům (srov. LEBEDA 2009).

<sup>26</sup> Srov. derivační slovník (analyzátor) pro němčinu, který je dostupný na adrese <<http://www.canoo.net>>.

<sup>27</sup> Jsou to brněnská Ajka (<<http://nlp.fi.muni.cz/projekty/ajka/ajkacz.htm>>) a pražský systém Morfo (<<http://ufal.mff.cuni.cz/morfo>>) za akademické, komerční potom představují produkty firem Lingea (<<http://www.lingea.cz/nastroje.asp>>) a Eridanus (<<http://empyreum.com/ces/partner/eridanus/>>).

ani způsob datové reprezentace, ani lingvistickou koncepci (ve formě tzv. tagsetu<sup>28</sup>), dokonce ani výchozí jazyková data je dána historicky, personálně i světonázorově. Přesto existují alespoň snahy, byť jen mezi akademickými týmy, některé idiosynkratické překážky většího sblížení odstranit. J. Hlaváčová tuto iniciativu popisuje takto:

*„Při popisu jednotlivých morfologických kategorií budeme vycházet ze závěrů tzv. Morfologického konkláve (dále jen Konkláve), které se sešlo 21.–23. října roku 2005, aby definovalo jednotlivé morfologické kategorie (...). Jeho závěry však nikdy nebyly dovedeny ke zdárnému konci. Složení Morfologického konkláve bylo (podle abecedy, bez titulů): Jan Hajič, Jaroslava Hlaváčová, Karel Oliva, Klára Osolobě, Karel Pala a Vladimír Petkevič. Dokument, který na jeho podkladě vznikl, nebyl nikdy publikován.“* (HLAVÁČOVÁ 2009a: 2)

Na celém snažení je nejpozoruhodnější hned jeho pojmenování. Konkláve, tedy „přísně uzavřené a oddělené shromáždění kardinálů volících papeže“ nebo „místnost, kde se volí papež“ (NASCS), symptomaticky dotváří obraz vzniku a rozvoje české počítačové morfologie, tedy něčeho, co je hermetické, vzniká vskrytu a o čem je dostupné jen minimum informací. Velmi ilustrativní příklad pravého opaku synergie a interoperability. To ale není vše. Zvláštní zprávou o teorii a praxi veřejné podpory v české vědě je existence aktuálně běžícího grantového projektu, jehož hlavní řešitel pochází z pracoviště, kde vznikl jeden ze jmenovaných akademických analyzátorů. V anotaci k projektu<sup>29</sup> *Morfologická analýza a tagging s minimálními zdroji* můžeme číst např. toto: „*Přitom chceme dosáhnout úspěšnosti srovnatelné se systémy, které vznikly standardním způsobem, ale za zlomek ceny, času a s minimem lingvistických znalostí o daném jazyku.*“ [zvýraznil J. L.] Takový je kontext, do kterého vstupuje poznaná nutnost zabývat se vznikem nového, totiž morfematického analyzátoru<sup>30</sup>.

## 0.5 Metody, východiska a cíle této práce

Podstatou předkládané práce je návrh analyzátoru, který bude kromě morfematické schopnosti zvládnout také analýzu (a syntézu<sup>31</sup>) morfologickou (a pro futuro i analýzy další). Určujícím je zde základní poznatek, že hranice morfémů i jejich identifikace se při ohýbání nemění. Funkční delimitace jednotek lexému stojící v pozadí analýzy se tak umožňuje vyrovnat s důsledky dvojí artikulace (A. Martinet) aplikované mutatis mutandis na vztah lemmatu a jeho flektivního paradigmatu. Jedno členění se v případě sloves promítá aplikací konjugačního (a derivačního) pravidla do několika set (v závislosti na koncepci negace, verbálních substantiv atd.) analytických kombinací výrazových exponentů a souvisejících gramaticko-lexikálních významů. Podobně nosným postulátem je skutečnost, že hranice a identifikace jednotek zůstává zachována i při fonologické (fonetické) transkripci (pěst-i-tel-Ø × /pjest'-i-tel-Ø/). Přirozenou konsekvencí obou východisek je možnost modelovat morfematickou stavbu neobyčejně ústrojně a efektivně. Zvláště patrné je to při určování flektivních vlastností analyzovaného lexému. Slovo *bufet* je na grafematické úrovni buď sklonné (1) nebo nesklonné (2), tedy

*bufet 1: Nsg bufet, Gsg bufetu, ... a bufet 2: Nsg bufet, Gsg bufet, ...*

<sup>28</sup> Tj. systému značek kódujících nejčastěji gramaticko-lexikálních jazykové kategorie.

<sup>29</sup> Projekt č. GPP406/10/P328, podrobné informace viz *Informační systém výzkumu, experimentálního vývoje a inovací* (<<http://www.isvav.cz/>>).

<sup>30</sup> Není bez zajímavosti, že finančně minimalistický návrh projektu *Morfematická analýza češtiny* předložený v roce 2006 autorem této práce Grantové agentuře UK nakonec nebyl schválen. Text návrhu je spolu s oběma posudky publikován – sine ira et studio – zde jako příloha.

<sup>31</sup> Syntézu chápu v převládajícím pojetí jako proceduru odvození flektivního paradigmatu od libovolného slovního tvaru nebo slovního spojení (např. u numeralií).

Tak ostatně pracují dnešní morfologické analyzátory češtiny. V bilaterálním (grafematicko-fonologickém) a morfematickém pojetí je situace snazší. Tedy<sup>32</sup>

*bufet-Ø<sub>k1, k2</sub> = /bufet-Ø/<sub>k1</sub> /bufet-Ø/<sub>k2</sub>: Nsg bufet-Ø<sub>k1</sub> | bufet-Ø<sub>k2</sub> = /bufet-Ø/<sub>k1</sub> /bifě-Ø/<sub>k2</sub>,*

*Gsg bufet-u<sub>k1</sub> | bufet-u<sub>k2</sub> = /bufet-u/<sub>k1</sub> /bifě-Ø/<sub>k2</sub>*

Naznačený princip však a priori počítá se znalostí lexikálního významu jako s ontologickou entitou, která bývá zpravidla zachycena ve výkladových slovnících. Empirickému světu je však bližší přístup gnozeologický. Výslovnostní, resp. formálněmorfologická variace (/bufet/bifě/, resp. bufet, -u|-Ø) nepochybně nemusí být spojena s (nacionálním) významem po celou dobu existence lexému. Jednotlivé slovníky tak mohou ve vývojové perspektivě začít invariantním popisem (/bifě/), v dalším vydání počítat s dubletou a nakonec se hypoteticky opět dostat k invariantu (/bufet/). Rozdílný počet klasifikovaných významů i rozsah údajů o formálněmorfologických vlastnostech lexikální jednotky se navíc často slovník od slovníku liší. To ovšem nijak nemusí odrážet vývoj v jazyce jako takovém. Pokud by analyzátor byl postaven na jednom (příp. několika) lexikografických popisech, mohl by být reprezentativní vůči užití jazyka zaznamenanému v (psaných) korpusech? Za idealistické představy o možnostech a rychlosti práce lexikografů třeba i mohl. Do jaké míry by ale byl reprezentativní vůči (slovotvorné a morfologické) potencialitě jazyka? Jinými slovy, jak zajistit, aby byl analyzátor skutečně na konkrétních popisech nezávislý a zároveň jich mohl v plné míře využívat? Pokusím se odpovědět příkladem.

Ve viditelném českém internetu se nevyskytuje slovo *Praštěně*, morfematicky členitelné jako *Pra-štěn-Ø-ě*. Dalo by se uvažovat o tom, že tak majitel pojmenuje svou restauraci s důmyslnou aluzí výrobce piva *Pra-zdroj-e* a expresivního pojmenování pivního sudu o objemu 30 litrů, kterému se říká *štěn-Ø-ě*. Že by autor při pojmenovacím aktu tvůrce mohl reflektovat i tvarovou homonymii NAsg s adverbiem (od prašt-i-t), ponechám stranou. Zařazení do deklinačního systému substantivní flexe je snadné podle fundující složky, proto *Pra-štěn-Ø-ě*, *Pra-štěn-et-ě* atd. Dá se však očekávat, že ani v případě obchodních úspěchů nikdy nebudou existovat dvě *Pra-štěn-at-a*. Hypotetický slovníkový popis by mohl zkráceně vypadat takto:

**Praštěně**, -ete s. (jen v sg.) *Název restaurace ve městě XY.*

Pokud by se na základě tohoto popisu někdy rozšiřoval slovník morfologického analyzátoru, bylo výsledkem ohýbání podle vzoru *kuře* buď s vynecháním plurálních tvarů, příp. s přegenerováváním i pro množné číslo. Do strojového slovníku by se v závislosti na formalismu dostal buď kmen *Praštěn-* nebo by se využil tzv. prefixový guesser (srov. HLAVÁČOVÁ 2009a), který by nepotřeboval žádnou aktualizaci slovníku, neboť by zkombinoval prefix *Pra-* a starý záznam pro apelativum *štěně*. V úzu je ovšem dost časté ponechávat obchodní a firemní názvy nesklonné. Zvláště krkolomně to zní ve sportovních zprávách, pokud nese příslušná (extra)liga název sponzora pojmenovaného víceslovně. Potom by bylo nezbytné počítat i s fakultativní indeklinabilitou, tedy např. *Budu obědvat v Praštěně × Z Praštěněte půjdem(e) na fotbal*. S tím si ovšem ani guesser (resp. disambiguace, tj. proces zjednoznačnění morfologické analýzy v kontextu) neporadí. Morfematický analyzátor by měl.

Jeho návrh tak bude postaven na těchto třech konstrukčních principech:

1. morfematická segmentace jako podmnožina morfologické analýzy
2. dualita grafické a fonologické dekompozice základem analýzy i syntézy
3. nezávislost na konkrétním jazykovém popisu a podpora systémové tvořivosti jazyka

Protože však česká lingvistika prokazuje jistý deficit v kontinuitě morfematického výzkumu, jsou cíle této práce poněkud širší. Chef-d'oeuvre oboru, *Retrogradní morfematický slovník češtiny* Eleonory Slavičkové, dosud nebyl zhodnocen v míře, která mu právem náleží.

<sup>32</sup> Morfematická hranice je naznačena spojovníkem (-), fonologickou transkripcí ohraničuje lomítko (/), varianty odděluje svislice (|) a dolní index označuje kvalifikátor.

Ve své podstatě však nejde o dílo samé, ale o podstatnou část jazykového systému, který představuje. Umožnila to především podrobná interpretace naznačených hranic segmentace a souvztažných jednotek, kterou jsem uskutečnil se záměrem odvodit modely vnitřní struktury slov platné pro jednotlivé slovní druhy. Kvantitativní zpracování vybraných jevů (fonemické složení, kombinatorické možnosti, slovnědruhovú distribuce spod.) bude demonstrováno jednak v systému (langue), zde představovaném RMSČ, jednak v jazykové praxi (parole), zde zastoupené korpusy ÚČNK.

Aby měl proponovaný návrh potřebnou interdisciplinární relevanci, budu se zabývat i aspekty, které mají podpořit koncepční rozhodnutí obvykle činěná ad hoc, bez znalosti hlubších korelací jevů a dlouhodobých konsekvencí. První kapitola je tak věnována otázkám, jaký je sémiotický status morfému, je-li skutečně každý morfém znakem a jaký je podíl jednoduchých jazykových znaků na utváření znaků komplexních. Obsahem druhé kapitoly je přehled poznatků z oblasti psycholingvistiky a kognitivních věd, které se především týkají organizace mentálního slovníku. Sledovat budu především podstatu některých principů fungování produkce a percepce, jež mají velmi blízko k postupům morfemické syntézy a analýzy. Třetí kapitola se obrací zpět do bližší i vzdálenější minulosti, tak aby mohla poodhalit význam pozapomenutého předterminologického období zkoumání morfemické stavby jazyka. Morfémem v kontextu stratifikační lingvistiky a jeho nejednoznačným začleňováním do různě vymezovaných rovin jazykového systému se zabývá kapitola čtvrtá, která je rovněž pokusem o shrnutí dosavadních poznatků z oblasti teoretické morfemistiky. Empirický morfemický výzkum měl dosud, zejm. v jednotlivých národních slovanských lingvistikách, nejčastěji lexikografickou podobu. Představení nejdůležitějších děl, s větším důrazem na české a slovenské práce, je předmětem páté kapitoly. K morfemické analýze dnes existují dva poněkud protichůdné přístupy, první manuální (lingvistický) a druhý automatický (programátorský), porovnání jejich předností i nevýhod zasvětim šestou kapitolu předkládané práce. Sedmá kapitola podává přehled kvantitativních charakteristik apelativní části morfemického systému. S využitím údajů čerpaných z největších českých synchronních korpusů nabízí i konfrontační pohled na autentický úzus psaných textů. Tématem osmé kapitoly je vlastní návrh efektivní počítačové reprezentace morfemické analýzy i syntézy. Devátá, závěrečná kapitola stručně poukazuje na aplikační potenciál morfemické dekompozice psané i mluvené podoby jazyka.

Teoretické zázemí této práce tvoří především morfemické koncepce E. Slavičkové, Miroslava Komárka, akademické Mluvnice češtiny, stejně tak jako poznatky slovenských badatelů soustředěných kolem Miloslavy Sokolové. Formální stránka algoritmického popisu bude založena na standardech informatického přístupu ke zpracování přirozeného jazyka.

# 1 Anizotropie jazykových jednotek v empirickém světě (sémiotické expozé)

## 1.0 Motto

*„A tak lingvistika pracuje neustále s pojmy vytvořenými gramatiky, o nichž nevíme, zda opravdu odpovídají konstitutivním faktorům jazykového systému. Jak to však máme zjistit? A jde-li o přeludy, jaké reality můžeme postavit proti nim?“ (SAUSSURE 1996: 136)*

## 1.1

Hráč, který na první zkoušce orchestru nerozpoznal dirigentův signál k nástupu, svůdník, jenž byl zatčen pro obtěžování, když si v bulharském hlavním městě Sofii špatně vyložil dívčino příkryvnutí, nebo zákazník v restauraci, který zaplatil prvnímu muži, který se objevil v bílé košili s motýlkem a s otevřenou peněženkou. Nebo příklady z jazyka. Pokud někdo tvrdí, že na těch slovech není ani zbla pravdy, nebo se omlouvá s vysvětlením, že má přehršle starostí, zkrátka *pech*, případně si stěžuje na kvalitu polokošil z tržnice slovy „To je ale peklo!“, pravděpodobně každý pochopí, o čem je řeč. Na druhou stranu, kolika rodilým mluvčím češtiny je dnes známo, že *zblo* pochází ze slova stéblo, že *přehršle* znamenalo původně málo, vlastně jen tolik, kolik se nabere do obou hrstí, výrazy *pech* a *peklo*, že mají dávný společný archisémém smůla, resp. dehet, nebo že polokošile není kompromis mezi tričkem a košilí, nýbrž oděv ke hře zvané pólo (srov. REJZEK 2009)? Jinými slovy, pokud nejobecnější definice znaku hovoří o zástupnosti něčeho za něco jiného (např. ČERNÝ – HOLEŠ 2004: 16, DOUBRAVOVÁ 2002: 28), do jaké míry lze u komplexního znaku předpokládat významovou samostatnost (diskrétnost) jeho složek a vice versa, tj. může mít komplexní znak nějaké další významové rysy neodvoditelné ze součtu sémantiky jeho komponent? A je množství znakových prvků složeného znaku dané objektivně, nebo epiricky?

Za jednu z obecně přijímaných vlastností jazykového systému je pokládána jeho anizotropie, tedy nesourodost, rozrůzněnost (srov. PALEK 1989: 20n, SOLNCEV 1981: 46). Aby mohly být jednotlivé prvky systému vzájemně rozlišitelné, musejí být diskrétní, diferencované co do svého složení (tvaru) a navíc hierarchizovatelné (hranice vyšších jednotek se shodují s hranicemi nižších). Lidský mozek tak sice dokáže vymyslet strojový kód počítače založený na kombinacích (lineárních posloupnostech) nul a jedniček, který je vysoce izotropní, sám by však tímto způsobem fungovat nemohl. Znamená však anizotropie jazykových jednotek per analogiam totéž co arbitrárnost jazykového znaku, totiž výchozí libovolnost, která se po zavedení, přijetí mění v závaznost? Jinak řečeno, existují v jazyce jednotky „stejně anizotropní“ pro všechny jeho uživatele? Případně ještě otázka další: pokud je možné v jazyce vymezit různé typy jednotek a jim odpovídající počet rovin popisu (viz kap. 4.1), mají všechny prvky téže roviny stejné vlastnosti, např. znakovou platnost? Naznačené okruhy otázek úzce souvisejí s povahou klíčové jednotky této práce, morfémem, a s kritérii, na jejichž základě je možné stanovit její hranice a odhadnout její kvantitativní proporce v systému jazyka. V novějších pracích (např. ESČ 2002, SOKOLOVÁ 2006 nebo ŠIŠKA 2005) je morfém per definitionem chápán jako bilaterální, tedy znaková jednotka. Při hledání odpovědí budu proto čerpat jak z pramenů obecnělingvistických, tak i sémiotických. Očekávané závěry mají napomoci praktickému cíli omezit v zásadách morfemické segmentace (kap. 6) subjektivní rozhodovací prvek na minimální, nevyhnutelnou úroveň.

## 1.2

Mezi tradiční výkladové postupy, kterými začíná úvodní část libovolného odborného textu, patří obvykle zpochybnění nociónálního obsahu základních termínů. Ani **sémiotika** (v Saussurově terminologii sémiologie) neboli nauka o významu a smyslu znaků (DOUBRAVOVÁ 2002: 9) není výjimkou: „Disagreements are widespread; they show that the term ‘sign’ is

*both vague and ambiguous. This disagreement extends to many other terms which are commonly used in describing sign-process.*“ (MORRIS: 1946: 3) Je proto logické, že ještě větší protichůdnost tvrzení platí v otázce, jestli je (každý) morfém znakovou jednotkou. Jisté explanační a definitorické průniky však přece jen existují, kromě toho výstižně formulovaný nesouhlas mnohdy napomůže tříbení pohledů více než pouhá parafráze s drobnými výhradami. **Znak** je nejčastěji definován jako „něco (a), co zastupuje něco jiného (b) vzhledem ke společné dispozici (c) sdílené mluvčím i adresátem“ (PALEK 1989: 15). Od neznaku jej odlišuje to, že je nositelem významu, který však povstává až v kontextu (HJELMSLEV 1972: 47, 49). „[I] části slova mohou být znaky (...).“ (HJELMSLEV 1971: 31) Platí při tom však známý postulát, že „[j]azykový znak nesjednocuje věc a jméno, ale pojem a akustický obraz“ (Saussure 1996: 96). Tedy alespoň v *bilaterální* znakové koncepci spojení označujícího neboli výrazu a označovaného neboli obsahu, která se na rozdíl od *unilaterální* (pouze výraz, vehikulum) a *sémiotické*, jinak též *triadické* (vehikulum, objekt a význam) v jazykovědě užívá nejběžněji. Proces, v němž něco funguje jako znak, se nazývá **semióza** (MORRIS 1997: 202), jinými slovy braní zřetele, nabývání (ztrácení) významu (DOUBRAVOVÁ 2002: 10)

Znak však není definován pouze funkčně, ale jsou mu připisovány také jisté ontologické vlastnosti: 1) arbitrárnost, 2) diskrétnost a obvykle lineárnost, 3) podíl na dvojí artikulaci, totiž původ v první a rozložitelnost na složky pomocí druhé (ČERMÁK 2001: 23). Konstitutivní složky znaku (Hjelmslevovou terminologií *figury*), tedy zejm. fonémy a hlásky, už znakovou platnost nemají, neboť postrádají významový korelát, mají pouze distinktivní funkci (ve vztahu ke znakovým jednotkám). Typologie znaků rozlišuje symptomy, signály, značky, odznaky, znamení nebo (podle Peirce) ikony, indexy a symboly (ČERNÝ – HOLEŠ 2004: 179–187). Znaky se účastí komunikační situace a podílejí se na utváření komunikačních funkcí (např. tři podle Bühlera, 1934 nebo šesti v Jakobsonově modelu, 1958) odvozovaných od jednotlivých participantů (původce, příjemce, předmět, sdělení, kód, komunikační kanál ad.) (DOUBRAVOVÁ 2002: 22, srov. OGDEN – RICHARDS 1997: 181). Fungovat (vstupovat do komunikace) může znak jen tehdy, „je-li jeho vnímatel kompetentní, tj. zná-li jeho význam a smysl nebo předpokládá-li alespoň, že znak nějaký smysl má nebo měl“ (DOUBRAVOVÁ 2002: 28). Podle některých je sémiotika samostatným oborem i metodou, uplatňovanou v jiných odvětvích (biologii, fyzice, matematice, právu, estetice atd.), existují však i názory, že je částí lingvistiky (BARTHES 1967: 62).

### 1.3

Nejasná je situace ohledně dynamiky znakových systémů. Znaky jsou někdy pokládány za „ustálené, a jsou tudíž takto základem pro synchronní komunikaci; musejí být proto pro všechny účastníky stejné“ (ČERMÁK 2001: 24). Jindy je zdůrazňována tendence znakových systémů „stále se dál vyvíjet“ (ČERNÝ – HOLEŠ 2004: 18n). Dokonce se mluví o relativní stabilitě (např. Morseova abeceda), příležitostných změnách (systém dopravního značení) a neustálé proměně, která má být vlastní přirozeným jazykům. Týkat by se měla jak formální, tak obsahové stránky s rychlostí diferencovanou podle jednotlivých rovin (ibid.). Zdá se, že např. znakový systém češtiny je více „pružný“ než „stabilní“, jak ostatně ukazují jazykové příklady z úvodu nebo oscilace významové (*přítel*, *data*, *vymazlený*, *sít'* aj.) i výrazové (*Fortuna extraliga*, *příjem prezidenta*, *Zubo[vo]vá* od Zubov ad.). Velmi blízko stojí problém motivovanosti, kdy je užitečné rozlišovat *motivovanost přímou* (ve vztahu k denotátu, předmětu), potom např. piktogramy patří mezi motivované a jazykové znaky s výjimkou onomatopoií mezi nemotivované (arbitrární), a *motivovanost relativní* (mezi jednotlivými znaky), která je vysoce individuální (ibid.). Jazykový, zejm. lexikální znak je tak zcela odlišně motivovaný pro rodilého a nerodilého mluvčího nebo spisovatele a etymologa. Pokud je však konstitutivním rysem znaku jeho význam, o jaký druh významu jde?

V jazyce existují jednotky se vztahem k jednotkám vyšším, nadřazeným, komplexnějším, ale také jednotky, které nemají „přímý vztah k žádné jednotce ostatních částí jazykového systému“ (PALEK 1989: 101n). Jsou to především slabiky (*trá-va, ha-dr, po-vo-dí*), tvořené podobně jako morfémy kombinací významuprostých fonémů, dalo by se však uvažovat i o kmenech (*-pracov- / -pracuj-, -umě- / -umí-*), majících jen přechodovou tvarotvornou funkci. Pozoruhodná je však nepoměrně větší dispozice mluvčích segmentovat psané nebo zvukové slovo na neznakové slabiky (určit jejich počet) než na znakové morfémy, tedy nejmenší jazykové jednotky, které nesou význam (ibid.: 106). Významová platnost morfému však patří mezi nejméně jasné konceptualizace jazyka. Argumentovat proti významu některých typů morfémů totiž zároveň znamená zpochybňovat jeho znakový status. Mluví se tak např. o tom, že je znakem neplnoprávným (Palek), neboť mu chybí objekt<sup>33</sup>, nebo „často oslabeným“ (ČERMÁK 2001: 23). Bývá také označován za abstraktní, čistě vztahovou jednotku (ERHART 1973: 61), minimální jazykový znak, který má vždy význam a/nebo funkci (ČERMÁK 2001: 122). V hierarchii morfémů se někdy kořenu připisuje lexikální význam, což je způsob zjednodušení přijatelný snad jen v případě absence slovotvorných afixů a při nulových (analogii vydělitelných) gramatických sufixech (afixech), např. *strom, on, mlád* atd. Podle Erharta je však (v indoevropských, semitohamitských ad. jazycích) kořen pouhou fikcí lingvistů a „jeho význam představuje abstrakci z významů lexémů, v nichž je obsažen“ (ERHART 1973: 66).

Asymetrický vztah mezi formou a obsahem (SKALIČKA 1935, PALEK 1989: 106n) vede jednak ke vzniku homonymie (lexikální znak), resp. synkretičnosti<sup>34</sup> (gramatický znak), jednak synonymie, příp. významové afinity. Předpokládá se tak např. samostatný morfém čísla, pádu, jmenného rodu apod., který však může mít buď paralelní (redundantní) lexikální vyjádření (*optimální × nej-optimáln-ějš-í* /superlativ/, *člověk × lid-é* /číslo/). Z toho důvodu se ve významové rovině znaku rozlišuje *sémém* (význam lexikálního morfému) a *gramém* (význam gramatického morfému), příp. *sémantém* (obsah kořenového morfému) a *formantém* (obsah derivačních morfémů) (ERHART 2001: 67). Výrazovou stránku znaku pak reprezentují příslušné morfy (EŠČ 2002: 273). Tím je ale znejasněna strukturní distinkce systémové jednotky morfému (langue) a její komunikační realizace morfu (parole), které lze rovněž nahlížet bilaterálně. Velmi nezřetelný je také vztah mezi lexémem (slovem) a morf(ém)em, a to především u synsémantik<sup>35</sup>. Z morfematického úhlu pohledu je slovo „takové minimální uspořádání morfémů, kde pozice morfémů je vždy pevná“ (PALEK 1989: 124). Implicitně se tak předpokládá, že každé slovo sestává minimálně z jednoho morfému, flektivní potom ze dvou (byť někdy s nulovým exponentem). Kontradikční pojetí však operuje s neohebnými slovy jako „nemorfémovými“, „v jejichž stavbě se neurčují morfémy“ (SOLNCEV 1981: 52).

Jak ale naložit s pomezními případy, kdy neskloňná synsémantika variují za pomoci jiných morfémů (*na-te, vid'-te, to-t', ahoj-ky*), příp. reduplikace (*ne-ne, cha-cha-cha*) nebo kdy autosémantika využívají ve svých analytických formách neplnovýznamová slova (*se/si, by-0, by-ch(om), by-s, komu-s, včera-s, ředitele-s*), resp. transpozicí (konverzí<sup>36</sup>) přecházejí k synsémantikům (*nedbaje, myslím, stranou, okolo*)? Zachovává se původní morfematická

<sup>33</sup> To je ale i případ lexémů typu l-ový, énický, x-krát, teď aj.

<sup>34</sup> V koncepcích, které neuznávají nulový morfém (např. ROMPORTL 1970), je rezultující synkretičnost (v povrchové struktuře) neúnosně vysoká a nesystémová (srov. KOMÁREK 2006b: 128). Např. pro imperativní tvar *běž* je tak nutné všechny gramatické kategorie, které vyjadřuje, asociovat s jeho jediným morfémem namísto distribuce podle paralelních tvarů *běž-0-0 × běž-0-me, děl-ej-me*.

<sup>35</sup> Za synsémantika – kromě obvyklých prepozic, konjunkcí a partikulí – v této kapitole pokládám z praktického hlediska i interjekce (abych nemusel vždy uvádět synsémantika a interjekce), a to pro jejich společnou neohebnost a především nesamostatnost významu (závislost na kontextu), který je však roven funkci (např. význam *do háje, no nazdar, a sakra, propánajána* aj. je ve společné (komunikační) funkci exprese (zaklení), podobně vně kontextu nejasný význam *prásk* – funkce imitativní).

<sup>36</sup> Srov. BEDNAŘÍKOVÁ 2009.



delimitace (synsémantik), vzniká jeden nový morfém, nebo slovo zcela pozbývá morfematického statusu? Okruh otevřených otázek je však ještě širší. Pokud má mít každý znak význam nebo funkci (příp. obojí), jakou interpretaci přisoudit segmentu -iv- v adjektivu *šed-iv-ý* × *šed-ý* nebo -il- v *mal-il-inkatý*? Není to ani případ nulové povrchová realizace (morfu) reprezentované v hloubkové struktuře příslušným sémémem (např. řekl + Ø[/jest] × jsem) nebo gramémem (např. *řekl-Ø* × *řekl-a*), ani pouhý funktiv (např. a /parataxém/ či deadjektivní formant -ě /hypotaxém/ (ERHART 2001: 84)).

#### 1.4

V morfematických systematikách (MČ1: 177, SOKOLOVÁ 1999: 29) se pro části slovního tvaru, které mají jen výraz a nikoli význam, užívá souhrnný termín *submorf*, nikde už se ale neuvádí jejich znaková platnost. Do této sumární kategorie spadají jednotky, jež bývají jinak označovány podle funkce např. jako *interfixy* (konekty) a *infixy* (tematické a rozšiřující sufixy, reziduální a přehodnocené morfémy). Vedle sebe se tak ocitají segmenty se zjevnou funkcí (spojovací, systemizační aj.) a takové, u nichž žádné teleologické zdůvodnění možné není. Tedy přinejmenším prizmatem syntagmatických a paradigmatických vztahů mezi různými typy jednotek a souvisejících rovin popisu. Slabika ale jednotkou gramatiky není, a přesto má pro morfém zřetelnou důležitost (v souvislosti s přízvukem, suprasegmentálními jevy) (PALEK 1989: 116). Podle mého soudu je to právě projev komunikační potřeby zkracovat nebo prodlužovat délku fráze, přízvukového taktu apod. (*řek'* × *řekl* × *řeknul*), ať již z důvodů stylistických (*šedý* /2slab./ × *šedivý* /3slab./) nebo ekonomických, tj. pod vlivem principu minimálního úsilí (*neškrť*; *Ředitel's neviděl?* apod.).

Do značné míry se na tom podílí průhlednost, popisnost morfematické struktury, resp. její motivovanost v očích jednotlivých uživatelů jazyka. Ta je ovšem značně disparátní, neboť ovlivnitelná vzděláním a proměnlivá v čase. Nelze jistě předpokládat, že se např. značkovost výrazu *respiračtor* pro někoho změnil tím, že se setká (uloží si do svého mentálního slovníku<sup>37</sup>) s příbuznými slovy *respirační* a *respirovat*. Po usouvztažnění s výrazy *in-spirovat* (pův. v-dechovat), *a-spirovat* (při-dechovat), *spirituál* a třeba *spiritualita* je však pravděpodobnost docela vysoká. Nutno podotknout, že podobný postup používají autoři morfematicky členěných slovníků<sup>38</sup> (viz kap. 5), je to vlastně jediný plauzibilní pokus, jak se vyhnout subjektivnímu hodnocení. Intuitivní představa fungování celého systému primárního (u nového slova) i sekundárního (nalézání nových vztahů mezi již známými slovy) motivování má podobu genealogického stromu, v němž existují jak bohaté příbuzenské vazby, tak osamocené větve bez přímých potomků. Jak známo, někdo se o svůj rodokmen nezajímá, jiný má prozkoumány své genetické kořeny několik století zpět. I zde však může nastat situace, že je někdo mylně pokládán za potomka někoho jiného, než s kým ho pojí důležité sekvence genomu. V jazyce jde o riziko *lidové etymologie* (např. *anti-datovat* /proti-/ × *ante-datovat* /před-/), která vytváří vývojově neopodstatněná spojení zpravidla na základě blízkosti formy, příp. homonymie. Někdy může být falešnou indicií synchronně neprůhledná hlásková změna (*vůbec* od *v obec*), příp. snaha o ekonomizaci artikulační námahy (*pondělí* z *po neděli*, *trpaslík* z *tři a pět*, *padesát* z *pět* a *deset* atd.).

Co to ovšem znamená ze sémiotického a morfematického hlediska? Přinejmenším to, že by mohlo být přínosné převést meritorní segmenty ze sumární kategorie *submorf* do nově zavedené třídy *stylému*<sup>39</sup>. Nejde však jen o nové označení, ale především o možnost funkčně stratifikovat některé dosud „prázdné“ segmenty. Důležitější je však odpověď na otázku

<sup>37</sup> Viz následující kapitolu, zejm. 2.4.

<sup>38</sup> I když s tím rozdílem, že se jazykovému povědomí napomáhá vyžíváním všech dostupných výkladových a etymologických slovníků.

<sup>39</sup> Podle EŠČ je *stylém* „Výrazový prostředek z kterékoli roviny gramatické a z kterékoli oblasti lexikální, který je nositelem stylistického příznaku (...).“

znakové platnosti submorf(ém)ů a morfematické segmentovatelnosti synsémantik. Pokud je obecně přijímána např. existence morfému čísla, tedy gramému, který sdílí morf (vehikulum) zpravidla společně s pádovým a rodovým morfémem (gramémem) – jako např. flektivní sufix *-a* v subst. *brambor-a* /sg + N + fem./ –, i když jde zjevně o pouhý význam (či je ovšem forma?), potom by neměl být problém přiznat status morfému, a tedy znaku také morfu, který se s jinými dělí o obsah (např. téma (morf) *-a-* ve slově *roz-děl-a-t* podílející se spolu s prefixem na morfému (gramému) vidu). Analogická asymetrie formy (více než jedna) a významu (jediný) bývá někdy označována jako *cirkumfix* (příp. asociované /Komárek/ nebo rozštěpené /Palek/ morfémy) a její uplatnění se spatřuje jak v lexikonu (typ *ná-břež-i*), tak v gramatice (perifrastické futurum *bud-* + inf., kondicionál *by-* + *-l-* aj.) (ČERMÁK 2008a). Per definitionem se mluví o „nespojité konfiguraci morfémů“, která je jako celek „funkčně ekvivalentní“ (jednomu) morfému (ibid: 78). Jde tedy o 1 sémém/gramém (význam) a 2 morfy (formy). S čím je ale např. *bud-u* vés-*t* ekvivalentní? Se syntetickou formě *po-ved-u*? Potíž tkví zejm. v neurčitosti toho, které vehikulum přesně odpovídá morfému (gramému) futura. Nepochybně *bud-* a *po-*, morfém infinitivu *-t* však velmi nejistě (*budu* / *musím* / *je mi* vést apod.). Tato nejistota ostatně vedla v některých morfologických směrech k opuštění morfému jako základní jednotky ve prospěch slova (srov. BEDNAŘIKOVÁ 2010). V indoevropských, semitohamitských, bantuských aj. jazycích se totiž vyskytují případy, kdy dvěma (příp. i třem) gramémům odpovídá jeden společný morfém (ERHART 1973: 82). Tedy 2 / více gramémů a 1 morf(ém) (vehikulum). Odpověď na otázku, kolik přesně morfémů zkoumané slovo (tvar) má, tak sotva může být složitější.

Pokud však platí, že závazným rysem znaku je jeho materiálnost (SOLNCEV 1981: 86) a že morfém má vždy význam a/nebo funkci (ČERMÁK 2001: 122), neměl by být nematerializovaný gramém (čísla, pádu, času apod.) pokládán za samostatný morfém, nýbrž pouze za distinktivní rys asociovatelný buď s nějakou parciální složkou konkrétní manifestace lexému, nebo s lexikální jednotkou jako celkem ((*přá*+/*pl.*/)-*tel-Ø* n. *přá-tel*-(*Ø*+/*pl.*/) n. (*přá-tel-Ø*)+/*pl.*/, zde však vč. gramatické reduplikace: *-Ø* = /*G pl. mask.*/ + /*pl.*/). Netýká se to však tzv. nulového morfému, který má nemateriální vehikulum systémově zakotvené, tzn. analogicky odvoditelné v rámci stejného slovního druhu, rodu, čísla i pádu (*Gpl fem. žen-Ø* podle *růž-i*, *písn-i*, *kost-i*). Kategoriální gramém tak nemá morfematickou platnost, to jej však nezabavuje statusu znaku, i když v poněkud neobvyklém chápání unilaterality (jen význam, tj. pouze příslušnost k abstraktní substanční třídě, nikoli výraz). Cirkumfixy a tzv. prázdné morfémy naopak materiální podobu mají a s doplněním nového typu *stylému* (*šed-iv-ý*, *mal-il-inký* aj.) všechny disponují nějakou funkcí. Proto je lze podle definice pokládat za samostatné morfémy (a pochopitelně i znaky).

V případě výrazové asymetrie, kdy jeden význam/funkce přísluší více výrazovým exponentům, je nezbytné počítat s existencí *vázaných forem* (již BLOOMFIELD 1933/1965). Ty se objevují jak v oblasti lexikální (např. morfém *se/si* u sloves), tak i gramatické (např. kondicionálové *by-*). Ze své podstaty je každý morfém nesamostatný (realizovatelný až prostřednictvím celého slova), a to i v případě tzv. kořených slov (*strom* × *strom-u*). Vázané formy však implikují obligatornost společného výskytu doprovázeného stejným významem. Týká se to i lexikálních morfémů, a tedy zejm. slovních kořenů (tzv. jaderné a rozšiřující elementy u Slavičkové (RMSČ: 30)). Tak např. morfémy (alomorfy<sup>40</sup>) *s-* a *-(e)k-/-(e)č-* jsou vázanými, rozšiřujícími morfémy ve vztahu k radixu, jadernému morfému *věd* ve slovech jako [*s-věd-(e)k*]-*Ø*, (*o*)-[*s-věd-č*]-(*i-t*) ad. Vázanost je ovšem oboustranná, protože samostatný (volný) radix *věd* (resp. lexémy, v nichž je obsažen) znamená např. ve slovech *s-věd-om-í* nebo *věd-ec-k-ý* zcela něco jiného. Prozatím jde tedy o tři morfémy s jediným (zde lexikálním) významem (o hloubkových a povrchových interpretacích viz dále). Stejným

<sup>40</sup> Distinkci morf (realizace morfému) – alomorf (poziční varianta) v této kapitole pro jednoduchost vypouštím (viz kap. 4).

způsobem je vyložitelná i morfematická struktura víceslovných lexikálních jednotek (horem dolem, mandelinka bramborová, na vysoké noze atp.), ne tedy jen analytické tvary jednoslovných lexémů. Přechodovou fází nejlépe reflektuje (ortograficky dubletní) psaní dohromady a zvlášť (*na slepo* × *naslepo*), u něhož sémantika nepochybně nevychází z prostého součtu významů jednotlivých složek. Tímto způsobem však splývají jednotky dvou rovin (viz kap. 4), a sice lexikální a morfematické. V příkladu *horem dolem* se objevují dva kořenové morfémy, dva morfematické znaky s jedním společným denotátem ‘jak je to jen možné’ a zároveň jeden komplexní lexikální znak s jediným víceslovným označujícím a jedním označovaným. Podobné by to bylo i v případě *na slepo* (2 morfémy /radix/, 1 lexém) / *naslepo* (2 morfémy /prefix a radix/, 1 lexém).

## 1.5

Tím se organicky dostávám k morfematickému složení synsémantik. Důležité je zde rozlišit neplnovýznamová slova primární (např. *na*, *alespoň*, *mňam*) a sekundární (např. *podívej*, *sbohem*, *přinejhorším*, *abyste*, *začátkem*), vzniklé z jiných slovních druhů. Obvykle se synsémantika (zejm. ta primární) nečlení a předpokládá se status lexikálního morfu nebo komplexního slovního tvaru (KOMÁREK 2006c), u sekundárních se (např. ve slovnících) hranice segmentů podle původního slovního druhu sice naznačí, avšak toto rozdělení je interpretováno jako „formálně segmentovaný kořenový morfém“ (SOKOLOVÁ 1999: 23). O kompozitech se v souvislosti s neplnovýznamovými slovy neuvažuje. Radikální přístup nepokládá synsémantika za morfémy vůbec (SOLNCEV 1981: 52). Ani jeden způsob se mi však nezdá příliš vhodný. Ponechám-li různé cíle morfematické analýzy (edukační, vědecky explanační, praktickolexikografický aj.) stranou, nenacházím sémiotické (a obecnělingvistické) zdůvodnění, proč k synsémantikům přistupovat jinak než k autosémantikům.

Všechny potenciální morfémy mají materiální povahu, zpravidla vázaným způsobem disponují nějakou (nejčastěji syntaktickou) funkcí, a tak jediné, co je od autosémantik odlišuje, je zachování původního výrazu za současné změny významu, resp. funkce. Výchozí členění např. deverbativního citoslovce *po-dív-ej-Ø* zůstává, ztrácí se však všechny slovesné gramémy (hypoteticky snad s výjimkou osoby a čísla) a slovo jako celek (jeho složky vázaně) dostává novou funkci (zde kontaktní prostředek). Navíc by měly být neutralizovány všechny nulové konstrukty (téma, osobní koncovka atd.), jejichž vysvětlení by nyní nebylo udržitelné. Kromě toho je, myslím, nutné počítat i s původním kořenem, který zde však získává rozšiřující elementy (vázané morfémy). Jejich identita se však částečně také mění. Derivační prefix *po-* zůstává povrchově slovotvorným morfémem<sup>41</sup>, hloubkově jde však o vázanou část kořene *dív*, někdejší téma *-ej* se nyní povrchově stává slovotvorným sufixem, hloubkově naopak vázaným elementem radixu. V povrchové struktuře tak zůstává (až na nulové exponenty) přehodnocené členění původní, hloubkově (významově) vzniká nový (složený) kořenový morfém (3 vehikula, 1 význam).

Rozlišení hloubkové a povrchové struktury je ovšem nezbytné i pro autosémantika. Např. ve slově *čtyř-s-přež-í* zůstává morfém *-s-* původním derivačním prefixem jen povrchově, hloubkově se mění na slovotvorný morfém<sup>42</sup>, pozičně charakterizovatelný jako interfix. Sémioticky má tento morfém 1 vehikulum a 2 obsahy/funkce, citoslovce *po-dív-ej*

<sup>41</sup> Pro počítačovou reprezentaci je důležité, aby výsledkem (slovnědruhově, příp. jinak) neparаметrizovaného dotazu na prefix *po-* byla všechna lemmata a tvary, v nichž se s libovolnou funkcí vyskytuje. To platí pochopitelně i pro radix *dív*, kde by však mělo být rozlišeno, že je v tomto případě distribučně vázaný (jako v případě *věd* a *s-věd-[e]k/[e]č*).

<sup>42</sup> Bez této diferenciacie by nebylo možné evidovat vývojové trasy (přehodnocení) např. slovnědruhových transpozic přímo na morfémech (formantech), jichž se bezprostředně týkají (žah-a-t /-a- = téma/ > žah-a-dl-o /-a- = slovotv. sufix/.

potom 3 formy (vehikula) s jedním kořenovým významem ('kontakto­vý prostředek'), takto shodným s významem lexému *podívej*; *po-* a *-ej* mají navíc ještě každý po jednom distinktivním rysu (gramému). Nesložené slovo (nekompozitum) tak má sice vždy jeden kořen (u synsémantik roven lexému<sup>43</sup>), ten se však může skládat z několika vázaných morfémů. Platí to především pro synsémantika sekundární (*s-boh-em*, *při-nej-hor-š-ím*, *a-by-s-te*, *za-čá-t-k-em*), ze systematického hlediska je však výhodné segmentovat i primární (např. *a-le-s-poň*). U flektivních slovních druhů předpokládám kromě radixu vždy minimálně jeden další deklinační nebo konjugační morfém, tedy i u indeklinabilií (např. subst. *tabu-Ø*, *prim-a*), kde případný nulový exponent váže příslušné gramémy pádu, čísla, rodu apod. Jinak by hrozilo nelogické „přetížení“ kořene rozdělu­jící u odvozenin oba členy fundačního vztahu ((*tabu*+/*NGDAVL*sg,pl/) × *tabu-ov*-(ý+/*Nsg mask. pozit.*/)). U synsémantik je jednomorfémová varianta přijatelná (*a*, *ba*, *haf*). Počet morfémů však není u jednotlivých slov nijak stabilní. Výchozí rozdělení je závislé na aktuálním stavu systému (v širokém smyslu povědomí). U flektivních slov dojde k delimitaci lexikálního znaku na dva morfematické, zjednodušeně na část ohebnou a neohebnou, která může (ihned nebo kdykoli později) připustit další přerozdělení. Buď směrem k většímu počtu segmentů (rozpoznání souvislosti s jinými prvky systému), nebo méně pravděpodobně k snížení množství segmentů, tzv. perintegraci (*ves-n-ic-e* > *ves-nic-e*, protože ~~*ves-n-í*~~ > *ves-nic-k-ý*). Podle dnešní úrovně poznání by tak zůstalo *alabastr-Ø*, ale dále by se segmentovalo *tr-pas-l-ík-Ø*<sup>44</sup>. Tento stav se však dále může měnit.

## 1.6

Jsem si velmi dobře vědom toho, že snaha určit hranice, významy a funkce jednotlivých morfémů je východiskem značně abstraktního modelování systému jazyka (viz výše uvedená citace o kořenu jako lingvistické fikci). Mnohdy je tak snazší vymezit množinu segmentů, které být vehikuly toho či onoho druhu významu nemohou, tedy postupovat cestou negativních rysů (-), než určit jeden konkrétní (+). Rezignovat a připisovat všechny hodnoty jednomu lexikálnímu znaku (celému řetězci morfémů) však zároveň znamená popřít zjevnou hierarchii jednotek (např. radix × postfix) i nezpochybnitelnou utvářenost morfologických forem. Zejm. pro potřeby počítačového zpracování je však i pouhá aproximace velmi nosná (desítky miliony tvarů, statisíce kmenů, tisíce morfémů, stovky strukturních vzorců atd.) Podobně svízelné je pracovat na úrovni morfémů s dvojicí pojmů autosémantický a synsémantický slovní druh. Co si např. počít s kořenovým morfémem *baf*, pokud se může uplatnit jak interjekčně (neplnovýznamově) v *baf!*, tak verbálně (plnovýznamově) v *baf-nou-t*.

Ostatně, již delší dobu je přesvědčivě empiricky prokázáno (pro němčinu, ale platí to i obecně), že radix není slovnědruhově vyhraněn (wortartgebunden). Odpovědnost za slovnědruhovou diferenciaci skutečně, jak uvádí H. Bergenholtz, stojí na flexivních a slovotvorných morfémech: „*Ein Kernmorph hat die Möglichkeit, in allen Wortarten repräsentiert zu werden. Dem Morph kann man deshalb auch keine Wortart zuschreiben (es repräsentiert ja ein nicht wortartgebundenes Element). Es sind im Wort enthaltenen grammatischen Morphe, die die Wortart angeben (...). Nur Flexionsmorpheme repräsentieren wortartgebundene Morpheme.*“ (BERGENHOLTZ 1976: 66) I v tomto případě se tak ukazuje, že je více než výhodné nebudovat počítačový model na pojetích spjatých s konkrétní teorií či interpretací uložené do rychle zastarávajícího (tištěného) slovníku. Stanovení hranic morfematické segmentace, stejně jako identifikace delimitovaných entit se vyznačuje zvlášť vysokou dynamikou. Tomu musí být uzpůsoben i formální aparát zajišťující efektivní

<sup>43</sup> Nulový kořen předpokládám jen v naprosto ojedinělém případě morfologické formy (slovního tvaru) *Ø-Ø-ho* nebo *Ø-Ø-mu* (A, resp. D sg. zájm. on), a to analogicky k tvarům *n-ě-ho/n-ě-j*, resp. *n-ě-mu*.

<sup>44</sup> RMSČ má však ještě „*trpas-l-ík*“ a Sokolová – Moško – Šimon – Benko 1999 „*trpasl-ík-Ø*“.

počítačovou reprezentaci těchto procesů. Sémiotický pohled na věc, myslím, poskytl pro další kapitoly (zejm. 6 a 8) mnoho cenných podnětů.

### 1.7 Shrnutí

Kapitola promýšlí sémiotické konsekvence interpretování morfému jako jazykového znaku. Prizmatem bilaterálního pojetí předpokládajícího existenci výrazové a obsahové stránky morfému aplikuje základní definitorické vlastnosti znaku (zejm. materiálnost a obligatorní přítomnost významu a/nebo funkce) na speciální typy morfémů, u nichž se projevuje asymetrie formy a významu. Zabývá se tak především morfémy tzv. prázdnými (*šed-iv-ý*), u nichž se běžně žádný význam/funkce neshledává (1 výraz/vehikulum a žádný obsah), morfémy nulovými (*řekl-Ø*), které jsou však vymezitelné analogií (žádné vehikulum a minimálně 1 význam/funkce), a cirkumfixy neboli morfémy asociovanými (*ná-břež-í*, (*řek*)-*l-i by-Ø*), kde minimálně jednomu významu odpovídá více forem. U submorfů, jak jsou prázdné morfémy někdy nazývány, je navrhována interpretace některých typů jako *stylémů* (coby projev stylizace a ekonomizace), čímž dojde k podstatnému snížení počtu pouze výrazových morfémů. Podobný efekt má zavedení distinkce *volné* (jaderné) a *vázané* (rozšiřující) morfémy (*s-* a *-k/-č-* v [*s-věd-č*]-it), jejíž explikativní síla nachází ve spojitosti s rozlišováním *hloubkové* a *povrchové* identifikace morfémů (1 vehikulum a více obsahů) své uplatnění také u sysémantik a interjekcí (interj. [*po-dív-ej*] × verb. *po-dív-ej-Ø*) a víceslovných lexikálních jednotek (zde s rozlišováním znaků-lexémů a znaků-morfémů /*na slepo* = 2 radixy, 1 lexém/).

Zkoumání charakteru anizotropie jazykových jednotek ukázalo problematičnost přímých spojení mezi významem/funkcí a jeho bezprostředním vehikulem. To dává za pravdu kritikům (funkční) morfologie, kteří namísto morfému počítají raději se slovem / morfologickou formou jako základní jednotkou, která je nositelem všech vyjadřovaných významů. Stanoviskem této práce je, zejm. při zohlednění potřeb počítačového zpracování, počítat i nadále s existencí významových komponent, tj. gramémů (sémémů), které však nemají znakovou platnost, a proto tvoří pouhé sémantické distinktivní rysy jiných znaků (morfémů), příp. jejich komplexů. U sysémantik a interjekcí (vlastnost příslušející celému slovnímu druhu /prepozic, konjunkcí, partikulí a interjekcí/, resp. lexému, nikoli morfémům, z nichž se skládají) se tak modelově předpokládá existence jediného vázaného kořene (*[a-le-s-poň]*), který má minimálně jeden společný (sdílený) význam, ale několik forem (morfémů). Ty však mohou mít vlastní sémantické (neznakové) rysy odpovídající např. slovtvornému způsobu, jímž lexém vznikl. Autosémantika jsou vždy minimálně dvoumorfémová (dva znaky), a to i v případě indeklinabilii (*tabu-Ø*).

Výsledkem opakovaných úvah o morfematické motivovanosti lexikálních jednotek je postulát dynamické povahy morfematických hranic i variabilní identifikace vyčleněných segmentů. Na základě měněního se (individuálního i společenského) poznání vzájemných vztahů mezi lexémy může totiž např. docházet k větší specifikaci (zpodrobnění), a tedy nárůstu počtu (různých) segmentů a jejich interpretací v jazykovém systému, nebo naopak k abstrakci (přehodnocení, tzv. perintegraci), a proto i k snížení celkového množství entit v systému (*ves-n-ic-e* > *ves-nic-e*, protože ~~*ves-n-í*~~ > *ves-nic-k-ý*). Nesamostatnost jakéhokoli morfému, tedy i radixu, rezultuje v poznání, že např. z hlediska lexikálního významu není možné uvažovat o slovnědruhovém vyhranění kořenového morfému.

## 2 Morfém jako konstituent mentální reprezentace mysli

### 2.0 Motto

*„Máme-li rozumět české morfologii, je žádoucí znát způsob, jakým rodilí mluvčí češtiny dekomponují a analyzují slovní tvary. Na základě toho pak lze volit preferovaný způsob lingvistického popisu.“* (SMOLÍK 2009: 244)

### 2.1 Východiska

Mnoho lidských vynálezů a na nich založených postupů a výrobků se přiznaně i skrytě inspirovalo přírodními ději, principy a konstrukcemi. Záměr navrhnout efektivní počítačové zpracování morfematické analýzy (segmentace a identifikace) i syntézy (paradigmatická extrapolace analýzy a propojení slovníku na úrovni segmentů) bohužel s podobným postupem počítat nemůže. Lidský mozek je přístupný zkoumání pouze jako černá skříňka, u níž přibližně víme, co do ní vstupuje, a podobně je našim smysly rozeznatelné to, co z ní vychází. O tom, co probíhá uvnitř, máme jen nepřímá svědectví získaná vyhodnocením uskutečněných pokusů (viz např. ALTMANN 2005, PINKER 2008 nebo SCHWARZOVÁ 2009). Navzdory mnohoslibnému citátu ze záhlaví této kapitoly tak ve skutečnosti příliš opravdu prokazatelných poznatků k dispozici zatím není. O češtině prozatím téměř nic, o ostatních jazycích a obecně platných mechanismech produkce a percepce řeči v nedostatečné míře. Bez ohledu na to, že se touto problematikou zabývá hned několik vědních oborů (*psychologie, psycholingvistika, kognitivní lingvistika, neurologie* ad.). I z toho mála, co o těchto fenoménech dosud víme, je však možné fruktifikovat některé podněty, které mohou být využity v konceptu formální reprezentace morfematické struktury češtiny. Zaměřím se především na tato tři témata: *souvztažnost psané a mluvené formy jazyka* (2.2), *organizaci mentálního slovníku* (2.3) a *principy jeho fungování* (2.4). Nejprve podám vždy přehled vybraných poznatků a potom uvedu komentář reflektující nastíněný způsob uplatnění.

### 2.2 Souvztažnost psané a mluvené formy jazyka

Po celou dobu existence moderní lingvistiky, ať již její počátek ztotožníme se vznikem historicko-srovnávací jazykovědy v první třetině 19. století (srov. ČERNÝ 1996), nebo jej klademe blíže antice, či naopak současnosti, trvá sepětí psané a mluvené podoby jazyka. Jejich vzájemný vztah (rovnocennost obou, dominance jedné, primárnost vs. sekundárnost atp.) byl z různých metodologických pozic již mnohokrát podroben hlubokým analýzám, aniž by příznivci toho či onoho pojetí získali argumentační převahu. Čekání na nové světlo, které by do věci měly vnést psycholingvistické výzkumy, vyjadřuje neschopnost vyrovnat se s touto problematikou čistě lingvistickými prostředky. Základní otázka, „*kolikátého že řádu jsou vlastně písemné jazykové znaky: zda prvního, jak je přesvědčen J. Vachek, nebo druhého, jak hlásá R. Jakobson* (...)“, (ADAM 2009b) zůstává stále bez odpovědi. Směšování ontologického a fylogenetického přístupu, které je zde, zdá se, čím dál tím častější, však alespoň umožňuje dosahovat dílčích vítězství. Věhlas některých je – mnohdy neopodstatněně – udržován nekritickým obdivem, jenž vůči jeho původci získávají nové a nové generace lingvistů.

Známe je tak např. stanovisko Saussurovo, zdůrazňující subsidiární status psané formy jazyka: „*Jazyk a písmo jsou dva odlišné znakové systémy, přičemž druhý existuje výlučně proto, aby reprezentoval první.*“ (SAUSSURE 1996: 59) Z ontologického hlediska je opravdu nesporné, že nejprve si jedinec osvojuje verbální kód a teprve potom se učí jeho grafický pendant. Fylogeneze jazyka naopak ukazuje, jak se původní druhotnost písma (bez vztahu k jakémukoli konkrétnímu jazyku) postupně změnila (po nástupu knihtisku) v rovnocennost obou podob. Jejich další vývoj už má však asymetrickou dynamiku. Mluvená forma má již vrchol svého rozvoje (klasická rétorika) za sebou, zatímco písemná se dále překotně

„technologizuje“, tedy obohacuje na úkor svého protějšku. V tomto duchu se nesou argumenty W. Onga, které jsou založeny zejm. na těchto premisách: „*Bez psaní by gramotný rozum nemyslel a ani nemohl myslet tak, jak myslí, a to ani ve chvílích, kdy se zabývá psaním, ale za normálních okolností i tehdy, když se snaží uspořádat myšlenky do formy mluvené řeči.*“ (ONG 2006: 93) Je to dáno tím, že „[d]íky odstupu, který vzniká v důsledku psaní, se ve vyjadřování, které se zabývá bohatého, ale chaotického kontextu většiny ústních promluv, rozvíjí nový druh přesnosti. (...) Aby se člověk vyjádřil jasně bez gest, výrazů obličeje, intonace a reálného posluchače, musí obezřetně předvídat všechny možné významy nějakého konkrétního tvrzení pro čtenáře v jakékoli možné situaci a musí přinutit jazyk, aby fungoval tak, že bude jasný sám o sobě bez jakéhokoli existenciálního kontextu.“ (ibid.: 121)

Psaná podoba jazyka zároveň napomáhá směně poznatků a zkušeností na několikanásobně vyšší úrovni: „*Grafolekt [druh zavedeného psaného jazyka] v sobě nese stopy milionů myslí, které ho využívaly k tomu, aby s druhými vzájemně sdílely své vědomí. Vtloukly do něj obrovskou slovní zásobu, takže množství slov je v něm řádově mnohem vyšší, než by bylo možné v jakémkoli orálním jazyce.*“ (ibid.: 125) Distance, která v podobě písma vzniká mezi původcem a předmětem jeho zájmu, umožnila konstituování objektivitu jako nového typu poznávání světa. Tím byly položeny základy moderní vědy, jejíž *raison d'être* se odvíjí od spojení „přesného pozorování a přesného vyjadřování“ (ibid.: 145). Otázku „přímé, nebo naopak nepřímé cesty od písemných znaků k významu“ (ADAM 2009b) tak pokládám za víceméně řečnickou. Argumentuje-li totiž tazatel v jiném sporu (ADAM 2006–2007) tak, že nejvlastnější mateřštinou etnického Čecha je čeština bez přívlastků, a nikoli jediný její útvar, je význam v jazyce podobně komplexním<sup>45</sup> jevem.

Dopad připomenutých poznatků na počítačové zpracování morfematiky je v tomto případě nepřímý. Jako nevyhnutelné se jeví pracovat se spojitou datovou reprezentací (orto)grafickou i fonologickou, zde v podobě standardizované transkripce. Segmentace autentického zvukového záznamu (srov. MACHAČ – SKARNITZL 2009), resp. řečová syntéza založená na morfémech (srov. kap. 9), stojí zcela mimo možnosti nastíněného plánu.

### 2.3 Organizace mentálního slovníku

Mentální slovník (dále jen MS), tedy uspořádaný soubor slov „ve smyslu pojmů (uložených v paměti) mezi hláskovou podobou a významem“ a pravidel chápaných jako „operac[e] skládajíc[í] slova do kombinací, jejichž význam lze vykalkulovat z významů slov a způsobu jejich uspořádání“ (PINKER 2008: 379), se od dětství neustále vyvíjí tak, „aby umožňoval produkci i recepci textu“ (NEBESKÁ 1992: 77). Nejméně konfliktním poznatkem je odhad jeho předpokládaného rozsahu poměřovaný obsahem tištěných výkladových slovníků. Celkový objem slovní zásoby, kterou jednotlivec skutečně dokáže užít v praxi, napomáhají poznat např. slovníky pokud možno úplného díla nejvýznačnějších spisovatelů. Jejich rozsah se v závislosti na délce aktivní kariéry, šířce záběru a použité metodice pohybuje kolem hodnoty 100 tisíc slov (srov. např. ČERMÁK – CVRČEK 2009 a ČERMÁK 2007). Dispozice pasivní slovní zásoby vzdělaného dospělého člověka je pak předpokládána s horní hranicí kolem 120 až 250 tisíc slov (NEBESKÁ 1992: 74, PINKER 2008: 15). Co se týká struktury MS, shoda panuje snad jen v tom, že je to „ve skutečnosti soubor vysoce komplexních nervových okruhů“ (ALTMANN 2005: 84). Význam je potom výsledkem aktivace příslušných okruhů,

<sup>45</sup> „Psaní je vždycky tak trochu imitací hovoru, a v deníku tudíž předstírám, že mluvím sám se sebou. Tímto způsobem však ve skutečnosti se sebou nikdy nemluví a bez psaného nebo možná spíše tištěného textu bych ani nemohl. (...) Jakmile dojde k tomu, že se chirograficky vyvolaný cit pro přesnost a analytickou exaktnost zvnitřní, může tento cit samozřejmě zpětně působit na mluvenou řeč, což se ostatně i děje.“ (ONG 2006: 120–123)

které však od sebe nejde oddělit jako jednotlivé přístupové „kombinační zámky“ (ibid.: 92, 246).

Spíše než přijatelně pravděpodobná hypotéza o tom, jestli jsou slova v mysli uchovávána „vcelku“ nebo „v částech“, resp. zdali se dekompozice omezuje „na segmentaci morfematickou, nebo sestupuje[-li] až na rovinu fonologickou,“ (NEBESKÁ 1992: 76) převládá rozdělení badatelů podle příklonu k jednomu ze dvou nejrozšířenějších modelů kognice<sup>46</sup>: modularistického a holistického. Modularistická koncepce staví na předpokladu, že v mysli existují různé systémy pro různé funkce, které spolu na některých úkolech spolupracují (SCHWARZOVÁ 2009: 19). Holistické nebo také konekcionistické modely předpokládají „velké množství jednotek, které jsou vzájemně síťově propojeny“ (ibid.). Spoje mezi uzly jsou ohodnoceny tzv. váhami představujícími údaje potřebné pro aktivaci. Jednotlivé uzly jsou propojeny „budivými (excitačními)“ nebo „tlumícími (inhibičními)“ spoji (ibid.). Znalosti jsou zde kódovány na úrovni spojů, učení představuje modifikaci jejich vah. Tento model byl opakovaně počítačově implementován v podobě tzv. perceptronu a využíván např. pro učení se výjimkám v jazyce (srov. PINKER 2008). Neschopnost využívat např. proměnné však vede k nepřijatelnosti výsledků takových pokusů a k potřebám neustálých úprav modelu.

Nejnověji se tak začíná prosazovat představa hybridního modelu<sup>47</sup>, který např. v rámci teorie slov a pravidel předpokládá využití pravidlového modulu pro produkci pravidelných slovních tvarů a perceptronu (učicí se sítě) pro nepravidelné formy. Verifikace této hypotézy je hledána v závěrech pokusů, které potvrzují, že „lidé mají v paměti uloženy podobné formy, když používají nepravidelný tvar, ale používají-li pravidelný tvar, nemusí se na paměť vůbec obracet“ (ibid.: 180). Zajímavé je také zjištění, že je MS při osvojování (prvního) jazyka ve velké míře adaptabilní na specifika toho kterého jazyka. „[Angličtí, francouzští, ale i mluvčí jiných jazyků] dokážou využívat sebemenších detailů [např. koartikulaci] v řečovém vstupu k rozlišení jednotlivých slov ve svém mentálním slovníku.“ (ALTMANN 2005: 73) Zdá se tedy, že dospělí rodilí mluvčí různých jazyků postupují při zpracování svého jazyka odlišným způsobem. Tomu ostatně odpovídají obtíže spojené s pozdějším učením se cizímu jazyku, které naráží na nepřekonatelné překážky, zejm. v rovině fonetiky a fonologie (např. tzv. akcent).

Pro konstrukci formálního modelu „počítačové morfematiky“ nacházím v představených poznatcích především oporu pro intuitivní záměr rozdělit analýzu i syntézu na dvě části, pravidelnou a nepravidelnou. Konkrétní využití vidím však v trochu jiném duchu. Již Pinkerův postulát, podle něž „nepravidelné tvary přezívají díky memorování předávanému z generace na generaci; klesne-li frekvence výskytu, oslabuje se paměťový proces a využívají se pravidla (dochází ke zpravidelnění),“ (PINKER 2008: 189) nepokládám za zcela přijatelný. Existují totiž i jiné důvody, pro které se např. v češtině uchovávají přechodníky (škola a četba), kromě toho frekvence posiluje pozici jazykových prostředků v systému bez ohledu na jeho pravidelnost. Důležitá je i symetrie subsystému a analogie (vyrovnávání), proto se např. uzuálně šíří (s vysokou frekvencí) tvary jako *jí* a *ví* pro 3. os. pl. zapadající do subsystému prezntních tvarů téhož slovesa (*ví, vím, víš, víme, víte*), i když v paradigmatu a zejm. v derivátech převládá odvozovací základ (alomorf) *věd-* (*věděl, -a, -o, -i/-y, budu, -e, ... vědět; vědění, vědomí, věda* atd.). Předpokládám tak rozdělení jiné, v němž pravidly (algoritmem) budou vyjádřeny všechny známé tvary (i nepravidelné), „nepravidelný“ modul využívající učení (perceptron), zde tvarotvorného a slovotvorného modelu podle

<sup>46</sup> Podle NASCS (2005) ‘souhrn operací a pochodů, prostřednictvím kterých si člověk uvědomuje a poznává svět i sebe samého, kognitivní procesy, poznávání’.

<sup>47</sup> „Každá teorie, která navrhuje pro všechny operace jeden jediný mechanismus, nabízí zmetek a lidský mozek zákonitě podává lepší výkon.“ (PINKER 2008: 212)



frekvencí v systému a v úzu, bude mít potom na starosti veškerou sféru morfologické (morfematické) potenciality jazyka (neologismy, aktualizace atd.).

## 2.4 Mentální slovník a principy jeho fungování

Jak ukazují některá měření, efektivita fungování MS je neuvěřitelně vysoká: „*Mozku trvá asi čtvrtinu vteřiny, než najde slovní pojmenování nějakého předmětu a přibližně další čtvrtinu vteřiny trvá programování úst a jazyka, aby ho mohl člověk vyslovit.*“ (PINKER 2008: 15) Díky Williamu Marslen-Wilsonovi je zhruba od osmdesátých let dokonce známo, že k rozpoznání slyšeného slova dochází dřív, než ho mluvčí dosloví. Např. anglické slovo *slander* začíná být jednoznačně rozlišitelné na úrovni fonému /d/, protože je v ten okamžik vyloučena hypotéza pokračování v podobě slova *slant* (ALTMANN 2005: 83). Mezi další obzvláště pozoruhodné heuristiky, které lidská mysl využívá, patří tzv. priming (effect), tedy zkrácení času nutného pro zpracování informace o slovu a jeho významu, pokud je (vizuálně nebo auditivně) doprovázena pojmem významově blízkým (např. člun a loď, avšak také ruka a paruka) (ibid.: 19). Příslušná hesla MS totiž nejsou zpřístupňována, ale pouze aktivována, a to překvapivě podle formálních vztahů mezi slovy<sup>48</sup> včetně alternativních významů ambiguitních slov (ibid.: 93).

Prakticky vzato se tak např. při vyslovení slova *autorita* v mysli aktivuje nejen reprezentace spjatá s tímto výrazem, ale také reprezentace, které „*odpovídají položkám auto, autor a pravděpodobně také Rita*“ (ibid.: 17). Pro aktivaci určité části nervového okruhu jsou však relevantní jen některé druhy stimulů. Méně časté výrazy zvyšují reakční čas odpovědi oproti frekventovanějším slovům, i zde je ale možná akcelerace díky jinému viděnému nebo slyšenému slovu ze stejné oblasti významové. „*Mírná odchylka [výslovnosti, příp. hluk v pozadí] nemá za následek sebedestrukci lexikálního hesla, pouze snížení jeho aktivace, což znamená, že se tato aktivace může opět zvýšit, bude-li následující vstup s tímto heslem i nadále slučitelný.*“ (ibid.: 92) Také závěry získané empirickým zkoumáním využití pravidel a učící se neuronální sítě přinášejí podnětná zjištění. Praktická realizace předpokládá soutěžení mezi využitím obou přístupů (srov. např. VÁCLAVÍK 2007, SOJKA 2005). Odpověď na otázku, zdali jsou tvary téhož slova v MS reprezentovány samostatně, nebo se „*vytvářejí podle morfologických pravidel ve chvíli, kdy je toho zapotřebí,*“ (SMOLÍK 2009: 242) je tak v různých případech odlišná. Podle Pinkera je pravidelný tvar uložený v paměti „*užitečný pouze tehdy, je-li natolik silný, že ho lze rychle vyhledat*“ (PINKER 2008: 201). Pokud zvítězí informace uložená v paměti nad pravidlem, „*[i]nhibiční spojení směřuje z paměťového okénka do okénka pravidel a pomalu brzdí pravidlo na důkaz toho, že se našlo řešení, a pravidlo nakonec přestane účinkovat*“ (ibid.: 189).

Návrh formálního zpracování morfematiky z uvedených podnětů využít zejm. poznatky o aktivovanosti neuronových okruhů, v tomto případě všech potenciálních segmentací a identifikací v průběhu morfematické analýzy. U slov a tvarů, které jsou pokryty pravidly (korelace slovního tvaru, jeho složek a celého paradigmatu vč. analytických tvarů), je způsob analýzy analogický paměťovému hledání a postup po morfémech urychluje diskriminaci nevhodných kandidátů (např. aut-or-it-a × a-u-t-o-r-i-t-a, tzn. *auto-mat*). Při analýze neznámých tvarů, resp. jejich částí může postup mutatis mutandis odpovídat paralelnímu soutěžení mezi pamětí a pravidly. Rozdíl je však v tom, že nemusí jít o hledání celého tvaru, ale jen jeho části (např. *čist?o?nos?o-plen-a*). Principu primingu je možná využít po začlenění automatické morfematické analýzy do jiných, aplikovaných úloh, např. zjednoznačňování významu neboli tzv. *word sense disambiguation* (srov. kap. 9).

<sup>48</sup> Např. znalost gramatické kompatibility se při prvním výběru možných interpretací nevyužívá (viz ALTMANN 2005: 95).

## 2.5 Shrnutí

Zaměření kapitoly je povýtce utilitární. Přehled vybraných poznatků z oblasti kognitivních věd a teorie komunikace (vztah psanosti a mluvenosti, mentální lexikon, aktivovanost, priming, úloha pravidel v produkci a percepci) slouží jako inspirace pro návrh počítačového zpracování morfematické analýzy a syntézy. Z uvedených podnětů se jako obzvláště nosné zdají koncepty spojitého zpracování psané a mluvené (transkribované) podoby jazyka, binárního rozdělení na znalosti a pravidla nebo aktivovanosti na úrovni morfémů (např. *aut-or-it-a* × *a-u-t-o-r-i-t-a*, tzn. ~~*auto-mat-*~~).

### 3 Pojem morfému v dějinách (světové a) české lingvistiky

#### 3.1 Morfém jako termín (a pojem)

Příležitostné snahy o shrnutí procesálních proměn *pojmu* morfém v lingvistice bez národních přívlastků (např. ARONOFF – VOLPE 2006, BOSÁK – BUZÁSSYOVÁ 1985) se nejčastěji spokojují se sledováním vývojových změn *termínu* morfém, čímž se jako počátek morfematických dějin stanoví osmdesátá léta 19. století a za iniciačního autora prohlásí polský lingvista *Jan Baudouin de Courtenay* (1845–1929). Tento přístup se jeví o to racionálnější, že cum grano salis je „[p]odľa niektorých lingvistov (...) morféma azda najviac nejasný pojem v celej jazykovednej literatúre (...)“ (BOSÁK – BUZÁSSYOVÁ 1985: 15). Terminologické expoé pak prvotní vymezení – „**that part of a word which is endowed with psychological autonomy and is for the very same reason not further divisible**“ (cit. podle ARONOFF – VOLPE 2006) – dává do souvislosti s původní potřebou rozlišit fonetické a obsahové členění jazyka na bezprostřední konstituenty, tedy foném a morfém. Asymetrii dalšího vývoje obou termínů způsobilo především větší rozšíření fonému (u Saussura a dalších), resp. pozdější zájem o vyšší (komplexnější) roviny jazykové stavby (viz kap. 4.1) související se vznikem strukturalismu, zejm. jeho americké větve<sup>49</sup>. Nejcitovanější definice morfému tak pochází z erbovního díla moderní (strukturální) lingvistiky, kterým je Bloomfieldova kniha *Language* z roku 1933. Přesně zní takto: „**A linguistic form which bears no partial phonetic-semantic resemblance to any other form, is simple form or morpheme.**“ (BLOOMFIELD 1933/1965: 161)

Další definitorické divergence souvisely jak se zaměřením jednotlivých badatelů (např. deskriptivně-distribuční akcent u Harrise, Gleasona ad.), tak s ontologickým hodnocením sociálně-substančních vlastností jazyka. Známý je např. historický spor o „pomologické univerzálie“, v němž stáli proti sobě A. I. Smirnickij (1902–1954), zastánce existence reziduálních morfémů (např. čes. *mal-* v malina n. angl. *cran-* v cranberry) vydělitelných analogicky, a G. O. Vinokur, vyznavač pozitivistického rozlišení významu všech jednotek. První přístup v novějších definicích morfému převládl, jak ukazuje např. vymezení M. Dokulila (1912–2002), příslušníka druhé generace tzv. pražské školy, které pochází z konce padesátých let: „**nejmenší[.] dále neděliteln[á] část slova, která má přímý nebo nepřímý vztah k významu.** Konkrétně to znamená (...) existenci (...) též **takových morfémů, které (...) lze vydělit jen (...) na základě asociativní analýzy.**“ (VACHEK 2005: 110) [zvýraznil J. L.] Za více než půlstoletí dalšího vývoje oboru se konsenzuální jádro nejnovějších definic změnilo méně, než by se dalo čekat. Podstatné bylo rozlišení mezi morfémem a morfem<sup>50</sup> (viz 4.3.1.1 a 4.3.1.3), tedy jednotkou systémovou a realizační (parolovou), které se však v praxi uplatňuje značně nesystematicky (srov. např. STATČ 2009). Konsenzuální pojetí zaznamenané v *Encyklopedickém slovníku češtiny* charakterizuje morf(ém) takto:

<sup>49</sup> V Tezích (1929) Pražského lingvistického kroužku se o morfému (resp. o morfématu, tzn. neutrum) píše ještě v rámci *Základních úkolů synchronické fonologie*. Celá pasáž doslova zní: „4. Důležitým problémem lingvistiky, zvláště slovanské, je morfologické využití fonologických rozdílů (*morfofonologie* nebo zkráceně *morfonologie*). Složitě představy dvou nebo více fonémat schopných podle podmínek morfologické struktury slova se navzájem zaměňovat uvnitř téhož morfématu, totiž morfonémata, mají podstatnou úlohu v slovanských jazycích (např. v rušt. Je morfonéma *k/č* v *ruk/č* – *ruka, ručnój*). [nový odstavec – pozn. J. L.] Je nutno přesně synchronicky stanovit všechna morfonémata pro každý slovanský jazyk nebo dialekt a místo, které dané morfonéma může zaujímat uvnitř morfématu. [nový odstavec] Náléhavým problémem slavistiky je provést naznačený fonologický popis u všech slovanských jazyků a jejich dialektů.“ (TEZE 1929/1970: 39n) První definici morfému v rámci pražské školy, která by nebyla publikována ve fonologické práci, podal (podle VACHEK 2005) až B. Trnka ve svém příspěvku *Morfologické protiklady* (1958). Morfém v typologickém kontextu poprvé (opět podle Vachka) vymezil V. Skalička v roce 1935.

<sup>50</sup> Motivace je velmi podobná té, která vedla k dichotomii věta a výpověď.

## Morfém

- „nejmenší, na sémantické úrovni dále nedělitelná jednotka jazyka mající povahu znaku ve smyslu jednoty označujícího a označovaného“
- „relační jednotka určitelná na základě vztahu k sousedním jednotkám“

## Morf

- „výrazová jednotka jazyka izolovaná při segmentaci mluveného nebo psaného textu, řetězce fonémů a odpovídajícího grafického záznamu“
- „je vyjadřovacím prostředkem pro morfém“

(EŠČ 2002: 273)

Morfému jako termínu a vývoji jeho chápání jsou v lingvistické literatuře věnovány stovky a tisíce stran, o morfému jako pojmu se explicitně nepíše téměř nikde. Implicitní zmínky přinášejí jen velmi vzácné přehledy dějin tzv. předvědecké lingvistiky, tedy zhruba od zrodu klínového písma do vzniku historicko-srovnávací gramatiky počátkem 19. století. Z česky psaných pramenů se nabízejí např. Černého *Dějiny lingvistiky* (ČERNÝ 1996), Večerkovy *Jazyky v komparaci I* (VEČERKA 2008) nebo Koupilovi *Grammatykáři* (KOUPILOV 2007). Pátrání po genezi lidského poznávání vnitřní struktury jazyka je i v případě relativního dostatku sekundární literatury zatíženo nevyhnutelným množstvím aproximací, hypotéz a extrapolovaných důkazů. Očekávaná disproporce mezi potřebným úsilím a možným výsledkem, bohužel v neprospěch toho druhého, mě přivedla na myšlenku obrátit pozornost trochu jiným směrem. Světový kontext ponechám stranou a pokusím se na díle čtyř známých osobností z českých dějin (obecných i tzv. vnějšího vývoje jazyka) prokázat, že pojem morfému je přinejmenším o několik set let starší než jeho termín a že útlum morfematického výzkumu češtiny v posledních třiceti letech má v kontextu jeho více než třísetletého trvání význam pouhé epizody. Jejich jména jsou Jan Amos **Komenský** (1592–1670), Václav Jan **Rosa** (asi 1630–1689), Josef **Dobrovský** (1753–1829) a František Ladislav **Čelakovský** (1799–1852).

Z terminologického hlediska není bez zajímavosti přiblížení stavu, jaký panoval v oblasti českého názvosloví zhruba v polovině 19. století, tedy asi třicet let před zavedením morfému. Toto období je důležité zejm. tím, že vydáním *Německo-českého slovníku vědeckého názvosloví pro gymnasia a reálné školy* (1853) končí jedna velká etapa obrozenských snah o rozkvět české vědy užívající české terminologie. V této oblasti se tou dobou nejvýznaměji angažoval Pavel Josef **Šafařík** (1795–1861), který je mj. autorem i následujících dvou studií publikovaných v *Časopisu Českého museum*, jak se tehdy (ne)skloňoval jeho název. Jsou to *O tvoření slov zdvojováním kořene* (ŠAFAŘÍK 1846) a *Mluvozpytný rozbor čísloslova* (ŠAFAŘÍK 1848). Navazující ukázka přibližuje podstatu pojmů, které jsme dnes zvyklí latinskou terminologií označovat jako alomorfiu (radixu), derivační a gramatické sufixy, vokalické a konsonantické alternace nebo kmenotvorné přípony a submorfy.

„V jazyku slovanském, jakož vůbec ve všech indoevropských, vnikneme-li pilným a střízlivým rozbořením slov do skladu jeho hloub, nalézáme ve slovech něco většího rozsahu a rozeznatelného útvaru dvě veskrz rozdílné částky; jednu, kteráž v též, buď naskrze nezměněná, buď poněkud přijínačené způsobě v menším nebo větším počtu příbuzných slov, co jádro jejich, na němž hlavně poněti lpí, se vyskytá, tuto jmenujeme *kořenem* slova (*radix*); jinou, která k onéno zevnitř a sice od zadu přistupující, pouze k vyznačení potahy poněti čili postavení jeho do určité kategorie myšlení a mluvení slouží, a tuto nazýváme *zvukem* čili *slohem odvodícím* i *skloňujícím*, samo pak to určování a odlikování poněti pomocí těch takových zvukův *odvození* a *skloňování* (*formatio*, *flexio*), ku př. ve slovech *duji*, *duch*, *duchovníkův*, *duchovenství*, *dýchati*, *dýchavičností*, *dychtěti*, *dyšnými*, *duše*, *dušice*, *dušenka*, *duševného*, *dušenství*, *dušiti* a t. d. slabika *du-* čili *dy-* ke kořenu, ostatní pak všechny k odvození a skloňování náležejí.<sup>1)</sup>

Důmyslným vyšetřováním starších i novějších mluvozpytcův dostatečně dokázáno jest, že odvození a ohybování jest jen zvláštní způsob ústrojného skládání slov, t. že všechny ty přírostky,

přístavky a přísuvky, ježto nyní, odtrhnou-li se od kořene, pouhé zvukoplodé a smyslu prázdné zlomky býti se vidí, prvotně a původně byly slova, rovněž o sobě užívaná a svůj zvláštní význam mající, kteráž však ustavičným jich přibíráním k rozličným kořenům, k určování a odlikování ponětí v těchto zavřeného, vedle těchto o svou neodvislost a samostatnost přišla.

<sup>1)</sup> O tvoření a skloňování slov *vnitřním měněním kořene* zůmyslně zde zamlčeno, ježto to ku předležícímu předmětu nevyhnutelně nenáleží.“ (ŠAFARÍK 1846: 446)

Pokud uvážíme, že v té samé době pracoval např. F. L. Čelakovský (viz kap. 3.5) hned na několika projektech slovníků (českých, jinojazyčných i srovnávacích), které měly slovní zásobu organizovat podle kořenů, tedy vlastně slovotvorně (srov. kap. 5), pak se morfematický výzkum druhé poloviny 20. století, většinou usnadňovaný počítačovou technikou, dostává do zcela jiného světla.

### 3.2 Komenský (rekonstruování Kořenové a Janua)

Polyhistor a vizionář Jan Amos Komenský je autorem velkého díla, kam patří také *Janua linguarum reserata* (1631, česky jako *Dvěře jazyků otevřené* 1633). Jak uvádí Z. Hladká, „[o]bě jsou v podstatě věcně řazenými několikajazyčnými slovníky blízcími se svou strukturou i celkovým charakterem modernímu onomaziologicky uspořádanému thesauru anglosaského typu“ (HLADKÁ 2007: 174). Popularita tohoto díla byla ve své době nebývalá a dočkalo se mnoha vydání v různých jazycích. Jedna z českých verzí *Januy* byla tištěna v letech 1648 až 1849 a pro dějiny jazykovědné bohemistiky je zvlášť významná. Bohumil Ryba (1949) zjistil, že s jejím tiskem souvisí vydání dvou méně známých Komenského prací, konkrétně *Dictionarium Vestibulare Latino-Bohemicum* a *Fundamenta Grammaticae – Základové grammatiky*. Na samém konci *Základů* pak objevil poznámku o titulu **Kořenové jazyka českého**. Dílo je bohužel nezvěstné, ale jeho obrysy mohl Ryba rekonstruovat podle jiných prací, např. *Stamm-Wörter Der Deutschen Sprache*.

Obsah by měl být tvořen abecedně řazenými českými názvy z vestibulárního slovníku. Podobně jako v německém Dikcionáři mají být i v Kořenech graficky rozlišovány složeniny, ačkoliv ani v jednom díle není tento postup uplatněn důsledně. Ryba uvádí příklady jako *poCestný*, *poCtivý*, *poČasí*, *poHřížiti*, *poŘadí*, *poSmíšek*, které v zásadě korelují s principem uplatněným už ve *Fundamenta Grammaticae*, kde se objevuje např. *spoluHlasné*. Cílem prý bylo odlišit „kořená slova“ od „zdvojenin“ a „odvozenin“. Ryba se věnuje také úvahám o možné souvislosti (závislosti) rekonstruovaných Kořenů, Komenského latinsko-českého a česko-latinského *Thesauru*, zničeného při požáru v Lešně v roce 1656, a stejnojmenného slovníku Rosova. Je přesvědčen, že jedním z pramenů Rosova díla byl český text „pražské jezuitské edice“ z roku 1667 *Zlaté Dvěře Jazykův otevřené*. Za pravděpodobné pak pokládá, že druhým dokladem mohly být Komenského *Kořenové jazyka českého*.

J. Marvan vidí vliv Komenského na Rosu i v rovině koncepční, když mluví o tom, že „[p]ravzoru Komenského lze přičíst i myšlenku slovníku ‚skutečné češtiny‘, tj. slovníku výkladového a dokladového“ (MARVAN 2006: 147). K tomu připojuje i záměr sestavit slovotvorný slovník, „v němž by slova byla uspořádána do hnízd podle formy a významu kořeného morfému bez ohledu na různost předpon a variabilitu kořene (...)“ (ibid.). O duchovním vlivu slavného exulanta jsou přesvědčeni i jiní badatelé, letitá spekulace o přímém využití Komenského lexikálního materiálu Rosu ze naopak zdá být vyvrácena. O. Koupil s odvoláním na V. Petráčkovou připomíná, že Komenský Rosu nejspíš vůbec neznal, naopak Komenského práce byly v Rosově pražském působišti relativně dostupné. V každém případě je nejméně od poloviny 17. století prokázána existence zájmu o elementární analýzu českých slov, alespoň na úrovni kořene jako základu pro kompozita a deriváty.

### 3.3 Rosa (gramatika a slovník)

Václav Jan Rosa je dnes jen málokomu znám jako jurista, tedy svým občanským povoláním, mnohem větší ohlas vyvolávají jeho jazykovědné zájmy, gramatika a lexikografie, i když zdaleka nejproslulejší je jeho údajná záliba ve tvoření nových slov. Jak uvádí na pravou míru O. Koupil, Rosa nebyl ani zneuznaný stylist, ani provokatér, neoterizaci namísto toho „*pojímal jako prestižní činnost učenců, související s poznáváním světa*“ (KOUPILOV 2007: 296). Celé jeho dílo se ostatně vyznačuje velkou mírou organizovanosti a promyšlenosti, kdy např. teoretická koncepce svědčí o značném materiálovém výzkumu a zároveň slovník svým obsahem vyvrací předpoklad pouhé enumerace. Podle dojmu novodobých opisovačů a kolacionátorů, jak vysvětluje Koupil, konkrétní materiál Rosova *Thesauru* podává mocné svědectví o autorově augmentaci lexika podle poznatých zákonitostí a pravidel tvoření. To by bylo jen stěží myslitelné v případě řazení lexémů abecedně.

Z obou Rosových lingvistických počinů se tištěné podoby dočkala jen jeho latinsky psaná gramatika nazvaná *Čechořečnost* (1672), slovník zůstal za autora života pouze v rukopise a ani pozdější opisy, z nichž nejznámější – i díky nedávné elektronické edici<sup>51</sup> – je tzv. Neuberkův opis, nezachovaly dílo v úplnosti (schází celé písmeno M a část N). Již v době Josefa Jungmanna neměl manuskript původní celistvost, takže jeho využití jako významného zdroje pro Jungmannův vlastní, abecedně řazený slovník (1834–1839) nezprostředkovalo Rosovo lexikum čtenářům v potřebné šíři (zkratka dokladů *Ros.*). To je u silně hnízdujícího lexikonu, v němž jsou u každého významu uvedena příbuzná i odvozená slova (např. prefigovaná slovesa, ale i sufixalia), nemalá ztráta i pro jiná než uvedená písmena abecedy. Zhruba 30 až 40 tisíc hesel není uspořádáno jen na principu slovotvorných možností češtiny, Rosa využíval při koncipování slovníku ve velké míře skutečné i domnělé etymologické souvislosti mezi lexémy. Z. Hladká (2007) uvádí několik nepřípadných spojníc mezi slovy *skot* a *skok*, *láska* a *hláska*, *včela* a *čilost* nebo *ráj* a *rád*.

Princip výstavby konkrétního (makro)hnízda může přesvědčivě ukázat např. slovní zásoba zahrnutá pod kořenné slovo *Pálám* (‘*geft wlaſtnj ohně, když plamen z ſebe wymjta, to flowe pála*’). Objevuje se zde třeba *plamen*, *podpaľ*, *palivo*, verbální derivační řada *do-*, *na-*, *od-*, *po-*, *pro-*, *pře-pálati*, ale také alternací vzniklé *popel* nebo prefigované *zpáliti ſe* či *zápalčiwý*. Otázkou slovotvorné dynamiky se Rosa zabýval také teoreticky. Ve své mluvnici věnoval tak věnoval pozornost především přechylování a odvozování, uvedl na 45 sufixů s příklady. Ústřední princip budování slovníkových hnízd, totiž akcentování kořenu jako základu tvoření a v případě sloves tedy velký důraz na prefixaci, umožnilo Rosovi promýšlet otázku slovesného vidu. V linii gramatografického vývoje je právě s tímto tématem spojován nejčastěji. Kromě očekávané distinkce *perfektum* a *imperfekum* (tzn. slovotvorně *simplicia* vs. *composita*, prefixalia) počítal také s protikladem *singulare* a *frekventativum* (tj. primárně záležitost sufixace). Obě binární opozice se navíc mohly křížit (KOUPILOV 2007: 283).

Pokud by rozsah zpracované slovní zásoby byl o něco větší a moment zdůrazňování tvořivosti a potenciality systému naopak nebyl tak nápadný, navíc pochopitelně s potřebou inkorporovat nové lexikum, patřil by Rosa a jeho dílo bez většího váhání do přehledu moderních slovotvorných slovníků představených v páté kapitole. Tím naléhavěji, že jiný slovník takového typu pro češtinu stále neexistuje.

### 3.4 Dobrovský (Über den Ursprung, Bildsamkeit a mluvnice)

Jestliže Komenský a především Rosa dokázali docenit význam slovního kořene, byl to Josef Dobrovský se svou příznačnou akribií, kdo postavil jeho zkoumání na vědecký základ. Tématem se zabýval opakovaně a hloubka celkového záběru je *ceteris paribus* nedostižná. Jak ostatně podotýká V. Flajšhans, „*[t]ěch 1605 kořenových slabik [z Institutiones, 1822], z nichž v Lehrgebäude nalezneme asi 1000, jsou vskutku jádro i základ slovanštiny veškeré (...)*“

<sup>51</sup> Elektronická verze je přístupná na internetové adrese <<http://vokabular.ujc.cas.cz/>>.

(FLAJŠHANS 1929: 85). Dobrovský se však neomezoval jen na otázku kořenů. Vzdáleně inspirován pracemi F. C. Fuldy a J. Ch. Adelunga položil základ svých slovotvorných výkladů ve dvou studiích z konce 18. století, na něž pak navazoval v obou redakcích (1809 a 1819) své gramatiky. Byly to *Über den Ursprung und die Bildung der slawischen und insbesondere der böhmischen Sprache*, která vyšla roku 1791 jako úvod ke slovníku F. J. Tomsy *Vollständiges Wörterbuch der böhmisch – deustsch – lateinischen Sprache*, a *Die Bildsamkeit der slawischen Sprache an der Bildung der Substantive und Adjektive in der böhmischen Sprache dargestellt*, již napsal v roce 1799 a později spojil s vydáním svého vlastního slovníku z roku 1802, resp. 1821.

Dobrovský v nich podává obecnou koncepci tvoření slov, věnuje se problematice kořenů, přípon, ale také třeba počešťování cizích slov (*Über der Ursprung*) a zabývá se rozbořem konkrétního jazykového materiálu (*Bildsamkeit*). Kořeny např. člení na tři typy podle fonologického složení: 1) sestávající z jedné souhlásky nebo spojení konsonantu a vokálu, 2) tvořené dvěma souhláskami, 3) obsahující tři konsonanty, nejčastěji se samohláskou po prvních dvou. P. Hauser upozorňuje na Dobrovského pronikavé postřehy týkající se distribuce těchto typů: „*Ukazuje, že některé kořeny jsou podkladem stavby slov jistého slovního druhu, např. 1. třída kořenů slouží ke tvoření zájmen a partikulí.*“ (HAUSER 1959: 44) Nechce se až věřit, nakolik Dobrovského poznatky a závěry rezonují na vzdálenost téměř dvou set let, které je navzájem dělí, s obsahem nejnovějších prací věnovaných metodám morfemické analýzy. Při řešení otázky původnosti, tedy zdali je prius verbum nebo substantivum, využívá sémantické kritérium. Substantiva pojmenovávající děj i jeho výsledek považuje za primární, slovesa vyjadřující pouze děj chápe za sekundární, tedy odvozená. Uznává ovšem i výjimky: dštíti od dešť, zvoniti od zvon atd. (ibid.). Podobně připouští, že někdy může být obtížné rozhodnout, patří-li slovo mezi kořenná nebo odvozená.

Přímý citát z *Bildsamkeit* (v překladu B. Jedličky) je v mnohém modernější než některé pasáže z nejnovějších bohemistických prací<sup>52</sup>:

„Česká podstatná jména jsou buď kořenná slova nebo slova odvozená. Všechna pouhá kořenná slova jsou jednoslabičná; avšak ne každé jednoslabičné slovo je také kořeným slovem, na př. duch, čich, zrak, mlejn, stan, smrt, stav, nejsou pouhá kořenná slova, nýbrž podstatná jména vytvořená pomocí odvozovací hlásky ch, k, n, t, v a odvozená od sloves douti (duti), čítí, zřítí, mléti, státi, mřítí.“ (DOBROVSKÝ 1953: 291)

V Dobrovského terminologii zahrnuje *odvozování* jen sufixaci, *skládání* potom i prefixální tvoření. Již ve své první studii inventarizuje nejdůležitější formanty substantiv a adjektiv s vymezením jejich obecného významu a řadí je podle hlavního konsonantu. V mluvnicki místo toho využil fonetickou příbuznost hlásek se současným respektováním původu, frekventovanosti, alternačních vlastností atd. Kořeny třídil v *Bildsamkeit* zvlášť pro adjektiva a substantiva, v prvním vydání mluvnicky uvedl tabelárně radixy všechny slovníků druhů, pro druhé vydání je nahradil seznamem všech kořených slov. V oblasti dnešní morfonologie počítal s kvantitativními i kvalitativními vokalickými alternacemi (padati – pád, postiti – půst, blud – blouditi, pletu – plot, stud – stydětí), uznával i existenci nulového alternujícího fonému (mru – mřítí, sen – snítí, hrom – hřmítí, pneu – zápona). Podobně přistupoval i ke konsonantickým alternantám (k – č, h – ž, ch – š atd.). Ve své gramatice zpracoval i problematiku kompozit, jejichž třídění založil na respektování jejich výsledného slovního druhu a potom podle první části (HAUSER 1959).

Na Dobrovského pracích o tvoření slov (tedy vlastně tvarotvorbě i slovotvorbě) spočívá z dnešního hlediska vlastně jediný stín pochybností. Že je publikoval německy. Tím omezil jejich dopad prakticky jen na úzký okruh obrozenců. Naštěstí však v tomto ohledu našel pozorné čtenáře, kteří kongenialitu teorie a praxe nemohla přehlédnout. Václav Hanka

<sup>52</sup> Srov. např. seznamy kořenů a tzv. kořených slov uváděné F. Čermákem (ČERMÁK 2010: 196–198).

demokratizoval obsah Dobrovského gramatiky v několika českých vydáních, která místy dále doplňoval, Josef Jungmann zase uplatňoval slovotvorné zásady při rozšiřování lexika o nová, inovovaná a adaptovaná slova. Přestože nebývá Josef Dobrovský v slavistických pracích o morfematice zahrnut do seznamů použité literatury, je tam přítomen stále, třebaže implicitně.

### 3.5 Čelakovský (kartotéky a nerealizované plány)

František Ladislav Čelakovský patří nepochybně k nejpřekvapivějším jménům této kapitoly, ale je to právě on, kdo měl v oblasti morfematické analýzy největší ambice a kdo i přes různá omezení toho i – kvantitativně vzato – nejvíc dokázal. Odvrácenou stránkou jeho nezměrného pracovního nasazení, které nebylo v naprosté většině nijak institucionálně podpořeno, je nezasloužená indolence badatelské sféry. Pokud se dostává odborné reflexe jiné než spisovatelské, překladatelské a etnografické činnosti F. L. Čelakovského (srov. VEČERKA 2008: 81), je to překvapivě jeho posthumní *Čtení o srovnávací mluvnici slovanské* (1853), jehož vydání si ale autor ve skutečnosti nepřál. Nezkreslené informace o jeho lexikografických a gramatografických krocích a záměrech, byť jen s omezením na český jazyk, dnes nejsou téměř vůbec dostupné. Svě tu pohříchu vykonala i ideologická předpojatost vůči autorovi de facto jediného pramene, který je založen na původním, materiálovém výzkumu. Autorem studie *F. L. Čelakovský a slovanská jazykověda* je totiž „oborový normalizátor“ Jan Petr (PETR 1988). Zdá se, že mu nemá být odpuštěno ani sub specie aeternitatis, třebaže to byl např. on, kdo v roce 1989 prosadil novodobou reedici Jungmannova slovníku, připravenou k 150. výročí díla.

Novější zmínky o Čelakovského lexikografické práci vytvářejí dojem neuskutečněných aspirací, je tedy možné se dočíst, že „[p]okus o hnízdově sestavený slovník, který podnikl Čelakovský, **zůstal však jen v torzu**“ (MARVAN 2006: 146) [zvýraznil J. L.]. Zásadní stať o vývoji české lexikografie zase odbývá téma takto: „*František Ladislav Čelakovský měl jeho [J. Jungmanna] slovní zásobu převést do slovníku uspořádaného podle kořenů, tj. podle původní představy Josefa Dobrovského. **K realizaci tohoto plánu však nedošlo**, a to především z finančních důvodů.*“ (HLADKÁ 2007: 182) [zvýraznil J. L.]. Opakovaná marginalizace úsilí, které se údajně buď ani nevynaložilo nebo nevedlo k ničemu zaznamenání hodnému, může jen stěží – v již tak dost exkluzivním oboru archivního bádání – podnitit kohokoli k pouhé pochybnosti o správnosti vynášených soudů. J. Petr však podává zprávu o něčem úplně jiném, když uvádí :

1) „Český jazyk zpracovával Čelakovský lexikálně a etymologicky. Z jeho korespondence od r. 1837 (...) se dozvídáme, že Čelakovský zepsal na lístečích celý Jungmannův slovník a takto získaný bohatý lexikální materiál uspořádal podle kmenů a přípon. Chtěl jej vydat jako šestý díl Jungmannova slovníku. Tuto práci vcelku v rukopise dokončil (dochovala se v LA PNP, ale z blíže neznámých důvodů ji nevydal tiskem, ani ji nedovedl do konečné autorské podoby. V jeho rukopisné pozůstalosti se zachovala pouze rozsáhlá kartotéka tohoto slovníku českého jazyka na lístečích formátu 182 × 105 mm (...).“ (PETR 1988: 11)

2) „Vedle toho se dochovala rozsáhlá kartotéka českých slov řazených podle sufixů (-ina, -ín, -va, -ba, -ství atd.) a rozsáhlý soupis českých cizích slov, který nejspíše také byl určen k uveřejnění jako doplněk základní české slovní zásoby slovanského původu. (...) Chtěl ukázat starobylost češtiny a její derivační možnosti. (...) V r. 1837 začal Čelakovský připravovat etymologický slovník českého jazyka (se zřetelem k ostatním slovanským jazykům) na základě materiálu v Jungmannově slovníku (...). V práci Čelakovský pokračoval v následujících letech, avšak nedokončil ji a jeho etymologický slovník zůstal rukopisným torzem.“ (...) Žádná z těchto prací o české slovní zásobě nevyšla tiskem.“ (ibid. 24n)

3) „Zachoval se nám v rukopise část Čelakovského etymologického slovníku slovanských jazyků (...) Tento slovník zůstal v nedokončené podobě a po stránce odborné nepředstavuje žádnou mimořádnou hodnotu.“ (ibid. 29)



4) „Dochovala se také velmi rozsáhlá Čelakovského kartotéka připravovaného slovníku slovanských slovních kořenů (asi několik tisíc slov), doložených v místních a osobních jménech (...).“ (ibid. 30)

5) „Asi kolem r. 1827 se začal soustavně zabývat studiem polabštiny a postupně zpracovával její slovník podle kořenů a mluvnic (...). Kromě toho připravoval lužickosrbský slovník podle kořenů.“ (ibid. 11)

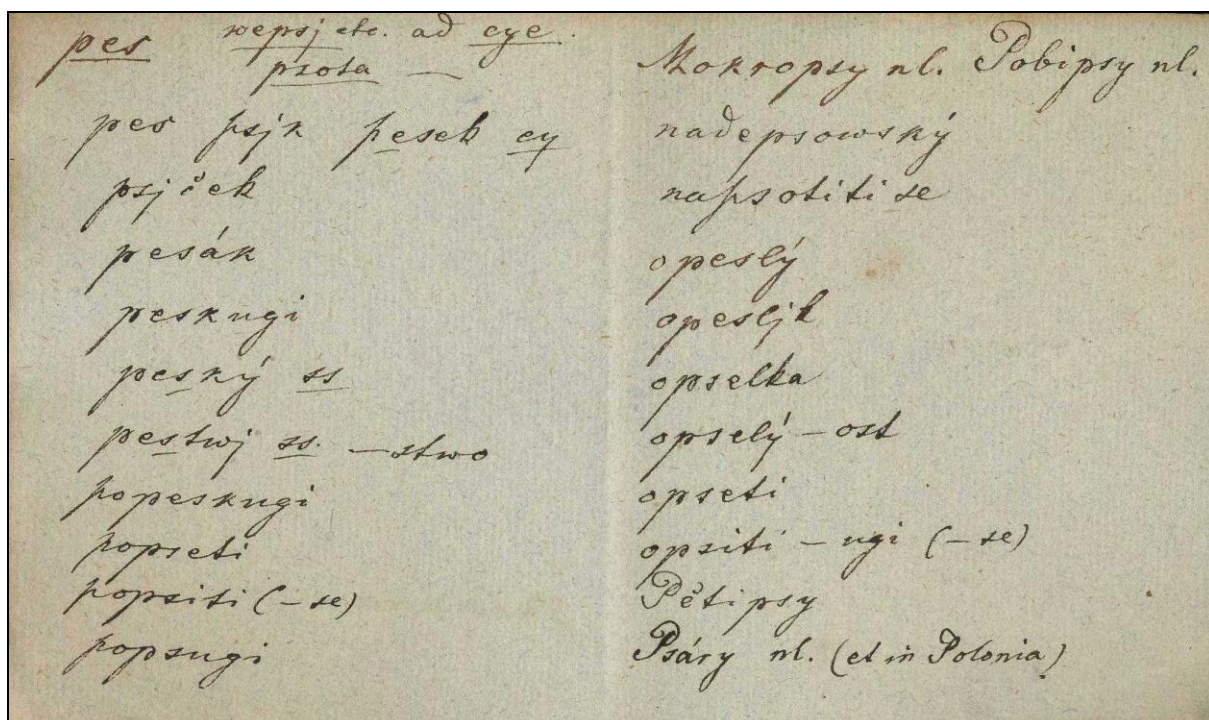
Uvedené citace nepokrývají celý Čelakovského slavistický záběr, vybral jsem jen ty pasáže, jež se týkají bohemistiky a jeho pracovní metody „podle kořenů“. Prakticky z ničeho však teď autor vyrostl v úspěšného lexikografa, jemuž osud nepřál dosáhnout zaslouženého uznání tím, že připravená díla vydal tiskem. Ani taková apoteóza však není na místě. Petrovo plaidoyer budí zbytečně velká očekávání, která provedené šetření v takové šíři nepotvrzuje. Pokud nepředpokládám, že se z osobního fondu F. L. Čelakovského, který je deponován v Literárním archivu Památníku národního písemnictví, za dvacet let od napsání Petrovy studie nic podstatného neztratilo, je potřeba umenšit především představy vyvolané sub 2 a 4. Rozsah zachovaného materiálu nebudí, až na kartotéky popsané sub 1 a jen částečně 2, dojem vysokého stádia dokončenosti, což ovšem nijak neoslazuje obdiv nad silou ducha schopného stranou běžných povinností dosáhnout takových výsledků.

Čelakovského kartotéka k slovní zásobě Jungmannova slovníku, uspořádaná podle kořenů a částečně podle slovtvorných formantů, jako jediná publikovaná bez větších zásahů umožňovala, nakonec se tak však nestalo. Důvodem, proč rukopis autorsky nepřipravil k vydání, mohlo být jeho stěhování do Vratislavi v roce 1840 nebo jenom nemožnost publikovat tuto práci s pomocí Matice, která v té době na podobné podniky neměla dost prostředků. V souvislosti s Jungmannovým dílem se mi naskytla příležitost zařídít digitalizaci současného stavu Čelakovského kartoték včetně doprovodných materiálů, takže mohu na základě přiměřeného poznání jejich obsahu konstatovat, že autor prvního známého slovanského slovníku slovtvorného, sestaveného podle kořenů, má jméno František Ladislav Čelakovský. V množství asi 21 tisíc naskenovaných polí (excerpčních lístků a dalších podkladů) připadá na českou slovní zásobu (mimo všeslovanské etymologické extenze) přibližně třetina. Největší část zaujímá hnízdové zpracování podle kořenů (5 797), lexikální jednotky sdružené podle zakončení (neúplný seznam sufixů i pouhých finál) představují objem o velikosti 693 excerpčních karet, soupis cizích slov (do Jungmannova slovníku nebyla z koncepčních důvodů až na drobné výjimky pojata) pokrývá plochu 98 lístků.

Kvantitativní odhady celkového množství zpracovaných lexikálních jednotek vycházejí z několika sond zjišťujících průměrný počet hesel na jednotlivých lístcích. Na základě prozkoumání cca 7 % obsahu hnízdových excerpt (412 lístků) vychází průměr 21,4 hesla na každém lístku. To by znamenalo objem slovní zásoby dosahující hodnoty 124 171 lexikálních jednotek. U inventářů slovtvorných prostředků předpokládám velikost 23 510 lexémů (693 lístků vynásobených průměrem 33,9 hesel), cizích slov je asi 3100. Přesnou podobu jednotlivých excerpčních karet konkretizují následující ukázky, obecně je možné říci, že obligatorní položku tvoří vždy záhlaví s jedním nebo více podtrženými kořenými slovy či přímo radixy (např. vě-ju; lep-ý, um), často s odkazem na jiné heslo, kořen apod. (např. gýt-ra cf. utr). Slovní zásoba byla pod příslušná záhlaví vypisována (a doplňována) zpravidla podle abecedy (hesláře Jungmannova slovníku), takže některá slovtvorná hnízda jsou rozdělena na několik lístků. Čelakovský nerozlišoval mezi apelativy a proprii, proto jsou např. pod záhlavím *pes* uvedena také (s latinskou zkratkou *nl.*) vlastní jména *Mokropsy*, *Pobipsy*, *Pětipsy* nebo *Psáry*. Při třídění slovní zásoby byly respektovány alternace kořene i další formální, méně často sémantické posuny. K heslovému slovu *druh* tak jsou přiřazeny např. lexémy *družnj*, *družný*, *-ost*, *-ně* (...) *přjdružný*, *-ost*, *mnohopodružný*, (...) *Mstidruh* no., *podruh* (...) *podružiti* (...) *podružný* (...) *přidruha*. U hesla *čtu* potom třeba *přepočtu*, *-ítám* (– *se*) (...) *přičtu*, *-ítám* (...) *rozčet*, *rozčetný* (...) *saučet*, *saučetný* (...) *sčjslný*, *sčjtanec* (...) *spočtowný*. Oproti Jungmannovu slovníku byla doplňována především vlastní jména, méně

často opomenutá či novější apelativa. Etymologické interpretace, které nezbytně souvisejí s přiřazováním lexémů k jednotlivým kořenům, nejsou pochopitelně neproblematické, není však možné poměřovat některé omyly dnešní úrovní poznání (např. kalous, kaloušek k uch apod.).

Přes všechny nedostatky zůstává Čelakovského slovníková kartotéka i po více než 150 letech od svého vzniku jediným českým pokusem o široce koncipovaný slovtvorný slovník češtiny uspořádaný podle kořenů. Jen podle odhadů musel autor ručně zpracovat, to mj. znamená rozhodnout o zařazení, přinejmenším 150 tisíc slov. Když navíc uvážíme, kolika dalšími jazyky, plány a úkoly se během pouhých 53 let svého života zabýval, mění se častá představa předních českých obrozenců jako poněkud umanutých podivínů v dojem nebývalé cílevědomosti, entuziasmu a pracovního nasazení, které v dnešní institucionalizované vědě nenachází mnoho paralel, pokud vůbec jaké.



Obr. 1: Ukázka Čelakovského kartoték (uspořádání podle kořenů)



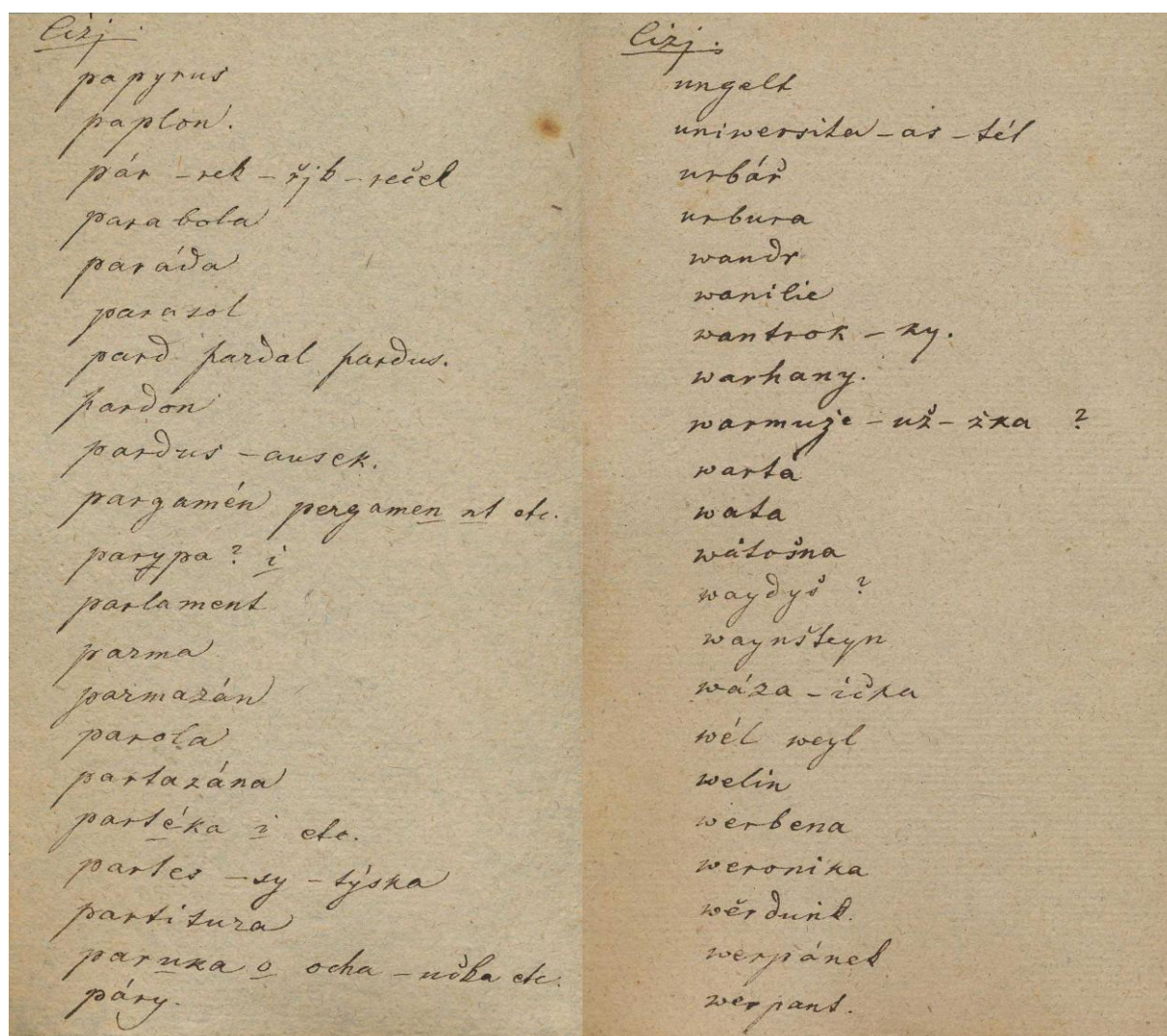
<u>mrn</u>	usmr titi - eugi
usmrlec	usmr sedlniti
usmr lina	usmrlný
usmrlný	x usmrswiti
usmrlo adv.	wšenesmrledlný
usmr titi	wymor dugi
usmr twěti	wymoriti - ugi
usmr twiti - ugi	wymru - jram
— elij	wymielj
umru umjram	žamor
	žamora
	žamoriti - ugi

Obr. 2: Ukázka Čelakovského kartoték (uspořádání podle kořenů)

<u>S m.</u>	<u>cn - in j.</u>
wěhlas	ručen
wědrus n et nl.	shlizen
winodus j - na	x shocen
wrus	spocen slo.
wlar - ásek - sejek.	slexen j - lina
wolopas	spären slo. - na
wonětras i	stágen - euka - neka.
wres a	x slawen
wšelisy nl.	x strassen
wjpros et - sa.	strázen - en
žákwas et - sa	studen
žánes únes	x studowen
žápas et - sa	sušären j j rna.
žáwěs et - sa	sypären j - rna.
žar	syren
aurus	x šiven
čjst	šloharen
črst no.	šufen ž
	šiken e
	x šichen j.
	šiken
	šmēn j ?

Obr. 3: Ukázka Čelakovského kartoték (uspořádání podle zakončení)





Obr. 4: Ukázka Čelakovského kartoték (soupis cizích slov)

### 3.6 Shrnutí

Úvod kapitoly připomíná disproporci pozornosti, jaké se v světové i české lingvistice dostává morfému jako termínu a morfému jako pojmu. První přístup se uplatňuje prakticky v každé větší práci o morfematice, resp. morfologii, druhý zůstává povětšinou opomíjen. Po nastínění hlavního směru definitorického vývoje je pozornost obrácena směrem k nereflektované tradici pojmoslovného dědictví minulosti, které se bezprostředně týká otázek vnitřní stavby slov. Trajektorie výkladu se dále soustředí na projevy zájmu o morfematické pojmy v dějinách české lingvistiky, z nichž vybírá čtyři osobnosti „předterminologického“ období, které je možné považovat za otce-zakladatele výzkumu české morfematiky. Po krátkém exkurzu, v němž je představen stav původní české terminologie zhruba 30 let před zavedením morfému jako termínu (J. Baudouin de Courtenay), se dílčí kapitoly věnují postupně J. A. Komenskému, V. J. Rosovi, J. Dobrovskému a F. L. Čelakovskému. U každého z nich se ve větší či menší míře objevují rysy moderních přístupů teoretického i praktického traktování morfematické problematiky, ať již užívají jakéhokoli pojmosloví. Tím se podařilo prokázat, že kontinuita morfematického zkoumání češtiny má mnohem hlubší kořeny, než se obecně předpokládá.

## 4 Morfematická rovina a její jednotky v systému jazyka

### 4.0

Obsahem této kapitoly je systemizace poznatků z oblasti formální morfologie, morfematiky, morfonologie a morfotaktiky. Nejen proto, že (s výjimkou Strakové (1985) a Čermáka (1990)) „[o]d Komárkových *Příspěvků* v českém prostředí mnoho pozornosti teoretické morfematice a morfonologii věnováno nebylo,“ (PLESKALOVÁ et al. 2007: 99) budu reflektovat i kontext slovakistický a obecnělingvistický. Specifickou problematikou (praktické) morfematické analýzy se bude zabývat samostatná kapitola 6.

### 4.1 Morfém fluktuující napříč abstraktními rovinami popisu

Lingvistika si je bez jakýchkoli pochyb jista předmětem svého zájmu, jímž je jazyk, méně již metodami jeho zkoumání. Nelze např. a priori rozhodnout, které přístupy jsou nadány větší explikativní silou, přesto však jsou některé úspěšnější v tom, že dlouhodobě plní očekávání svých uživatelů, což se sekundárně projevuje v množství přinejmenším plnohodnotných alternativ, jež přicházejí po nich. Jedním takovým je i jazykovědný strukturalismus, který jazyk nahlíží jako vysoce komplexní systém, tedy „soubor prvků a vztahů mezi nimi“ (SOLNCEV 1981: 20). Složité systémy sestávají z poměrně autonomních subsystémů, jež jsou rovněž definovatelné prostřednictvím svých prvků. Subsystémy bývají označovány také jako roviny nebo plány. Podle některých pojetí se zdá, jako by „vcházely jedna do druhé“ a při fungování jazyka se „uvádě[ly] do pohybu současně“ (ibid.: 73). Per definitionem je rovina „soubor relativně izotropních jednotek nestojících navzájem v hierarchických vztazích a mající hierarchické vztahy (ať už jako složky vyšší nebo nižší) k jiným jednotkám, jež také vytvářejí určitý soubor“ (ibid.: 71).

Počet rovin, jejichž existenci jednotliví badatelé předpokládají, a kritéria, která k jejich vymezení užívají, se někdy dost zásadně liší: „now the strata are based on the ‚sound (form) – meaning‘ dichotomy, now on various linguistic functions, now on different types of language means.“ (DANEŠ 2000: 13) Danešův přístup, založený na rozlišování rovin podle stejného stupně (znakové) komplexnosti a identity konstrukční funkce (ibid.: 15), počítá (včetně neznakové fonematické) s šesti rovinami: 1) fonematickou, 2) morfematickou, 3) lexikální, 4) syntaktickou a 5) textovou. P. Sgall vymezil, na principu „různosti kontextů“ a hierarchické diferenciaci forem a funkcí, v prvních teoretických zásadách tzv. *Funkčního generativního popisu* (SGALL 1967) celkem 5 rovin (viz obr. 5 níže), a to 1) fonetickou, 2) morfonologickou, 3) morfologickou, 4) větněčlenskou a 5) tektogramatickou (významovou). Nárůst komplexity jednotek (podle Hocketta (1967) vztahy C /skládá se z/ a R /reprezentuje/), se kterým zde počítá, je morfém „*reprezentován morfem, ale na morfologické rovině je složen ze sémat.*“ (SGALL 1967: 50) Dnes se ve Funkčním generativním popisu (FGP) počet rovin snížil o formálněsyntaktickou (4), takže se předpokládá přechod z tektogramatické přímo na rovinu morfematickou. V praktických aplikacích (např. v tzv. Pražském závislostním korpusu<sup>53</sup> (PZK)) však navzdory terminologii o žádnou morfematickou analýzu nejde (srov. HAJČOVÁ – PANEVOVÁ – SGALL 2002: 90). Morfematické (a nižší roviny) jsou totiž za „správně utvořený řetěz lexikálních a gramatických jednotek,“ (ESČ: 380) zodpovědné pouze teoreticky (implicitně), v praxi (rozboru autentických vět) se začíná až na rovině morfologické.

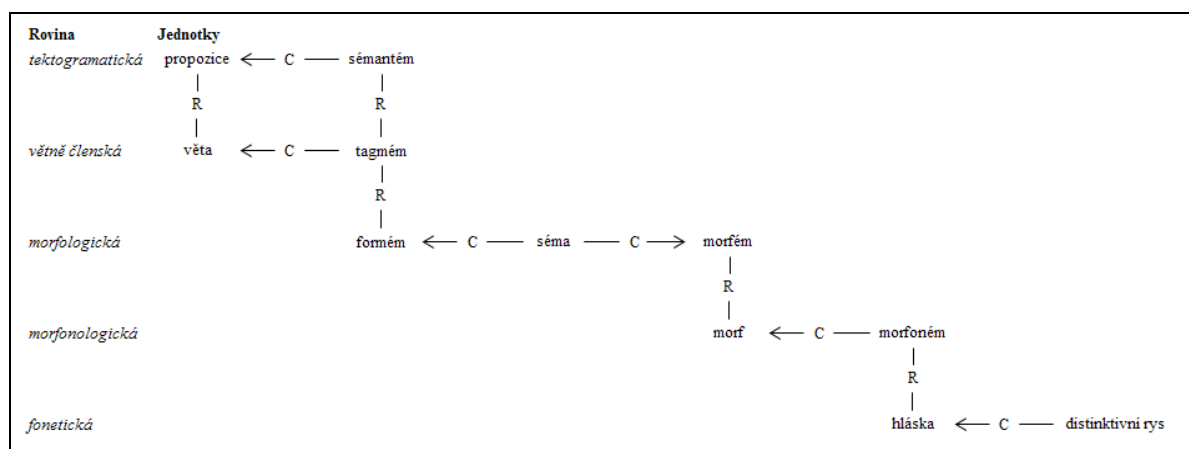
K jinému počtu rovin dospívá v rámci tzv. *Mean-Text theory* I. Mel'čuk (2006). Důsledným rozlišováním hloubkové (H) a povrchové (P) reprezentace jazyka zavádí hned sedm rovin: 1) sémantickou, syntaktické: 2) H, 3) P, morfologické: 4) H, 5) P a fonologické: 6) H, 7) P. V souladu s teoretickým konceptem se zde předpokládá opačný postup než např.

<sup>53</sup> <<http://ufal.mff.cuni.cz/pdt2.0/>>.

ve FGP – „from the most general to the most particular“ (ibid.: 12) –, proto je jako základní (nejnižší) brána rovina sémantická. Zatím nejvyššího počtu rovin (7) se dobral F. Čermák (2001), který v souladu se svým lexikologicko-frazeologickým zaměřením postuluje (obr. 6 níže) samostatné roviny apelativní, propriální nebo slovotvorné. Právě rozdělení morfologické roviny na flektivní a slovotvornou je již vzhledem k souvztažným jednotkám (morfému a afixu) značně konfliktní. Ani po terminologické záměně formantému (formantu) za afix by slovotvorná rovina nevyhovovala uvedeným kritériím (Danešovým ani Sgallovým) pro ustavení samostatné roviny. Žádná vyšší (komplexnější) jednotka se neskládá z afixů/formantů, neexistuje ani žádný hierarchický vztah formy a funkce. Jednotkami morfologické anebo lexikální roviny jsou tak výlučně morfémy.

Svébytný stratifikační model jazykového (obecnělingvistického) systému načrtl A. Erhart (1973), když důsledně rozlišil vztahy mezi plány (rovinami) sémantickým, lexikálně-gramatickým a zvukovým se zřetelem k prvkům, jednotkám (monémům) a vzorcům (tagmémům), jež je utvářejí. Tento přístup je velmi instruktivní, i když zavádí některé méně běžné termíny. Morfonem např. rozumí fonematickou reprezentaci morfému (spojení fonémů, jediný foném, příp. fonemickou nulu). Zřetelné rozlišení paradigmatických a syntagmatické vztahů vzhledem ke znakovým i neznakovým jednotkám a s usouvztažněním k sémantické dimenzi tak má zřejmě nejbližší k synkretické povaze jazyka. Obecné teoretické modely však zřídka bývají respektovány autory a rozvíjiteli specializovaných teoretických koncepcí. V dosud představených přístupech morfém (ať už byl definován jako jednotka roviny morfematické, morfologické nebo lexikálně-gramatické) nikdy nevstupoval do syntaktické (syntaktických) rovin(y) ani jako forma/prostředek, ani jako funkce.

Že tomu může být i jinak, ukazuje např. model tzv. distribuované morfologie hlásící se k přístupům generativní gramatiky, jehož základním principem je dekompozice morfémů na gramatické rysy (jako např. [Plurál] nebo [Kauzativnost]), které jsou uspořádány syntaktickými pravidly. „*Morfologické formy tak jen realizují určité části syntaktické struktury, forma a význam jsou do struktury distribuovány odděleně. (...) [O]ba typy afixů [derivační a flexivní] jsou jen formy, které realizují určité syntaktické rysy.*“ (ZÍKOVÁ 2007: 277) Přes takovéto a podobné vývojové tendence gramatických teorií, z nichž jsem připomínal např. postupné nahrazování morfému jako základní jednotky morfologie celou morfologickou formou, slovním tvarem (srov. BEDNAŘÍKOVÁ 2009), chápu pro potřeby své práce místo morfému v systému jazyka zcela tradičně. Následující přehled morfematických, morfonologických a morfofaktických pojmů počítá s rovinou jedinou, asi sice morfologickou, jejímiž jednotkami jsou morfémy (langue), resp. (alo)morfy (parole).



Obr. 5: Přehled rovin a jednotek (Sgall 1967: 56)

ZNAK				JEDNOTKA		LANGUE	PAROLE (realizace)
znak	gramatika	syntax	– nadvětná/textová/hypersyntax	promluva (propozice)	(textový typ)	text, promluva	
			– větná		(syntaxém)	(větný vzorec)	věta (textová)
		morfologie	– flektivní	morfém	morfém	morf (jako realizace)	
			– slovotvorná		afix	afix (jako realizace)	
			lexikon		(apelativa)	– slovní/kolokační	lexém – víceslovný
	(propria)	– (top)onomastická	\ – jednoslovný	lexém/slovo	lex (jako realizace)		
figura	fonologie		foném	foném (≠ hláska/zvuk)	fón (jako realizace)		
2. artik. : nesémant.							

Obr. 6: Přehled rovin a jednotek (Čermák 2001: 102n)

	PLÁN		
	sémantický	lexikálně-gramatický	zvukový
VZORCE (tagmémy)	vzorce hloub. strukt. vět (vzorce výpovědi)	vzorce povrch. strukt. (větné vzorce)	vzorce zvuk. struktury vět, slov a slabik
JEDNOTKY (monémy)	sémémy gramémy taxémy (?)	L-morfémy (lexémy) G-morfémy	fonémy intonace přízvuk předěl
PRVKY	sémant. složky (sém. dist. rysy)	morfony	zvuk. distinkt. rysy

Obr. 7: Model jazyka (Erhart 1973: 130)

## 4.2 Repetitorium poznatků (morfematika, formální morfologie, morfotaktika, morfonologie)

Následující přehled pojmů, termínů a konceptů, které jsou v různé míře sdíleny, avšak i zpochybňovány autory zabývajícími se formální morfologií, morfematikou, morfotaktikou a morfonologií, si klade hned několik cílů. V souladu se zaměřením přítomné práce na návrh formalizace a počítačového zpracování této lingvistické problematiky není smyslem ani sledovat vývoj<sup>54</sup> jednotlivých teoretických přístupů, ani podat jejich kritický komentář. Mým záměrem je naopak představit diferencovanost pohledů na přibližně stejnou oblast lingvistiky, jejichž intenzita se navíc – jak jsem připomenul v úvodu kapitoly – za posledních několik desetiletí prohlubuje spíše mimo oblast bohemistiky (srov. např. absenci pojmů sufixoid, postfixoid ad. v EŠČ (2002)). Tato terminologická i obsahová pluralita názorů, v souhrnu systematicky zaznamenávaná zkratkami pramenů, jednak demonstruje ekvivalenci teorií (např. morf a submorf vs. konektém a konekt), jednak odlišný zřetel k praktické aplikovanosti konceptů (např. cirkumfix / rozštěpený morfém traktovaný sémanticky (Čermák) a procesuálně (Slavičková)). Uvedený přehled tak umožňuje nahlédnout širší fenoménů, které je při formalizaci nutné brát do úvahy, a zároveň i (jmenovitě) rozeznat proporce pozornosti, jež jim byla dosud věnována.

<sup>54</sup> Viz kap. 3, v níž jsou naznačeny některé vývojové aspekty termínu a pojmu morfém, avšak rovněž bez ambicí podat vyčerpávající souhrn.

Ve snaze o koncizní, a proto v rámci možností přehledný tvar byla na minimální nutnou míru redukována problematika slovtvorná<sup>55</sup>, časový záběr zpravidla nepřekračuje zpětně hranici sedmdesátých let 20. století. Pro zachování maximální autentičnosti jednotlivých pojetí (a také z prostorových důvodů) jsem upustil od záměru do původních znění jakkoli autorsky vstupovat interpretací nebo parafrází. I přes jistou nesourodost výsledného znění se vědomě zříkám jiného komentáře, než jakým budu v 8. kapitole doprovázet jednotlivé části návrhu formálního zpracování. Zahrnuty byly především práce erbovní (KOMÁREK 2006c, SLAVÍČKOVÁ 1962, RMSČ 1975) a syntetizující (ESČ 2002, SOKOLOVÁ 1999, ČERMÁK 2001, PALEK 1989), výběrově jsem však vycházel i z děl obecnějších a také starších. Celkem jsem nakonec vycházel z patnácti zdrojů (viz níže), mezi nimiž stať Komárkova a dotčené výklady z *Encyklopedického slovníku češtiny* a prvního dílu *Mluvnice češtiny* byly excerpovány v úplnosti, a umožňují tak sledovat příslušné výklady relativně celistvě. Další prameny byly využívány k případnému rozšíření zahrnuté problematiky, ne již k opakování (potvrzování) výkladů nebo k dokládání méně významných odlišností. Terminologické paralely jsou naznačeny znakem rovnítka „=“ a/nebo zkratkou pramene uváděnou vždy ve složených závorkách „{ }“. Řazení jednotlivých definic a sentencí prioritizuje význam pramene a obecnost výkladu. Jazyk excerpt nebyl nijak upravován, místy bylo nezbytné provést pouze dílčí krácení a formální sjednocení. Příslušné prameny jsou identifikovatelné podle zkratky autora nebo díla a stránkového odkazu uvedeného za dvojtečkou.

Seznam zdrojů:

{Co} = Coates (2006); {Č} = Čermák (2001); {E} = ESČ (2002); {Er} = Erhart (1973); {H} Hockett (1967); {Ha} = Hay (2006); {G} Gleason (1969); {K} = Komárek (2006b); {M1} = Mluvnice češtiny 1 [MČ1] (1986); {P} = Palek (1989); {R} = RMSČ (1975); {Ro} = Romportl (1970); {S12, S15, S17} = Slavíčková (1962), (1965), (1967); {Sn} Solncev (1981); {So} = Sokolová (1999)

Obsah repetitoria:

<sup>55</sup> Základní pojmy a terminologii podávají jak díla klasická (DOKULIL (1962), DANEŠ – DOKULIL – KUCHAR (1967), DOKULIL – KUCHAR 1977, KUCHAR 1963), tak i novější (BOZDĚCHOVÁ (1994), MITTER (2006)).



## 4.3 MORFÉMIKA/MORFEMATIKA

### 4.3.1 Základní jednotky (nacionálně, třídění)

#### 4.3.1.1 morfém

##### 4.3.1.1.1 definice

##### 4.3.1.1.2 třídění

#### 4.3.1.2 submorfém = konektém {K}

##### 4.3.1.2.1 definice

##### 4.3.1.2.2 třídění

#### 4.3.1.3 morf

##### 4.3.1.3.1 definice

##### 4.3.1.3.2 třídění

#### 4.3.1.4 submorf = konekt {K}

##### 4.3.1.4.1 definice

##### 4.3.1.4.2 třídění

#### 4.3.1.5 alomorf = varianta morfému {M1}

##### 4.3.1.5.1 definice

#### 4.3.1.6 izomorfa

##### 4.3.1.6.1 definice

#### 4.3.1.7 morfoném

##### 4.3.1.7.1 definice

### 4.3.2 Morfologická paradigmatika

#### 4.3.2.1 slovní tvar = morfologická forma {E}

##### 4.3.2.1.1 složený/analytický/opisný/perifrastický tvar

#### 4.3.2.2 morfologická báze (slovního tvaru)

#### 4.3.2.3 slovotvorný základ

#### 4.3.2.4 tvarotvorný základ

#### 4.3.2.5 (tvaroslovný/tvarotvorný) kmen

##### 4.3.2.5.1 jednoduchý × odvozený

##### 4.3.2.5.2 infinitivní, přítomný, minulý

##### 4.3.2.5.3 pasivní, imperativní

#### 4.3.2.6 paradigma

##### 4.3.2.6.1 kategoriální × formální

##### 4.3.2.6.2 široké × úzké

##### 4.3.2.6.3 morfologické paradigma (slovních druhů)

#### 4.3.2.7 morfologické typy

#### 4.3.2.8 slovotvorné typy

#### 4.3.2.9 pád

#### 4.3.2.11 pádová polymorfie

#### 4.3.2.11 homomorfie

### 4.3.3 Funkční a konstrukční typologie morfémů

#### 4.3.3.1 třídění

##### 4.3.3.1.1 vázané × nevázané / volné

##### 4.3.3.1.2 samostatné

##### 4.3.3.1.3 funkční × prázdne

##### 4.3.3.1.4 jádrový

##### 4.3.3.1.5 hypermorfém (rozšířený)

##### 4.3.3.1.6 bimorf

#### 4.3.3.2 kořen(y)

##### 4.3.3.2.1 reziduální, zbytkový

##### 4.3.3.2.2 jaderné (a tzv. rozšiřující elementy)

##### 4.3.3.2.3 reziduální

#### 4.3.3.3 afixy

##### 4.3.3.3.1 gramatické/tvaroslovné/tvarotvorné

##### 4.3.3.3.1.1 definice

##### 4.3.3.3.1.2 třídění

##### 4.3.3.3.1.3 kumulativní morfémy

##### 4.3.3.3.2 slovotvorné

##### 4.3.3.3.3 prefixy/předpony

##### 4.3.3.3.4 prefixoid

##### 4.3.3.3.5 sufixy/přípony

##### 4.3.3.3.5.1 finální/koncovka

##### 4.3.3.3.5.2 sufixy funkční × prázdne

##### 4.3.3.3.5.3 tzv. složené sufixální morfémy

##### 4.3.3.3.5.3 spolymorfém (sufixu)

##### 4.3.3.3.5.4 derivační

##### 4.3.3.3.5.5 modifikační

##### 4.3.3.3.6 sufixoid

##### 4.3.3.3.7 postfixy

##### 4.3.3.3.8 postfixoid

##### 4.3.3.3.9 infixy a interfixy

##### 4.3.3.3.9.1 infix

##### 4.3.3.3.9.2 interfix

##### 4.3.3.3.10 raritní afix

##### 4.3.3.3.11 unikátní afix

#### 4.3.3.4 afixoidy

#### 4.3.3.5 lexikální morfémy

#### 4.3.3.6 alternační morfémy

#### 4.3.3.7 asociované (sdružené) morfémy = cirkumfixy

#### 4.3.3.8 rozštěpené morfémy {SI}

#### 4.3.3.8 konektémy (konekt – alokonekt) = submorfémy

##### {M1, So}

##### 4.3.3.8.1 redundantní, prázdne morfémy

##### 4.3.3.8.2 složený slovotvorný morfém

##### 4.3.3.8.3 kmenotvorný morfém

##### 4.3.3.9 kmenový konektém / kmenotvorný morfém {E}

### 4.3.4 Identifikace alomorfů

#### 4.3.4.1 alomorf (alokonekt) téhož morfému (podmínky):

#### 4.3.4.2 typy alomorfů

##### 4.3.4.2.1 nekomplementární

##### 4.3.4.2.2 komplementární

##### 4.3.4.2.2.1 paralelní

##### 4.3.4.2.2.2 synonymní

##### 4.3.4.2.2.3 supletivní

### 4.3.5 Souvztažné pojmy

#### 4.3.5.1 nulový morfém (konektém) / morfémová nula {E}

#### 4.3.5.2 morfémový/morfematický šev

#### 4.3.5.3 morfematický / morfémový uzel

#### 4.3.5.4 perintegrace / přerozdělení {E}

#### 4.3.5.5 deetymologizace

#### 4.3.5.6\*trunkace/trunkácia

#### 4.3.5.7 distribuce (libovolného morfému)

#### 4.3.5.8 centrum a periférie

##### 4.3.5.8.1 centrum

##### 4.3.5.8.2 periférie

#### 4.3.5.9 morfostylém

#### 4.3.5.10 odpovědnost a kompetence morfému

##### 4.3.5.10.1 odpovědnost morfému

##### 4.3.5.10.1 kompetence morfému

#### 4.3.5.11 fúze

## 4.4 MORFOTAKTIKA

### 4.4.1 definice

### 4.4.2 obecné poznámky

### 4.4.3 fonematické struktury morfů

#### 4.4.3.1 čeština

#### 4.4.3.2 slovenština

## 4.5 MORFONOLOGIE

### 4.5.1 Vývoj a obsah morfonologie

#### 4.5.1.1 definice

#### 4.5.1.2 morfoném

#### 4.5.1.3 morfofoném

### 4.5.1 Hláskové a morfologické alternace

#### 4.5.1.1 hláskové alternace

#### 4.5.1.2 morfonologické alternace / střídání fonémů

#### 4.5.1.3 alternační řada

##### 4.5.1.3.1 uspořádanost × neuspořádanost

##### 4.5.1.3.2 základní varianta

##### 4.5.1.3.3 princip většinové podoby

##### 4.5.1.3.4 fonémová nula / nulový morf

##### 4.5.1.3.5 alternativy korelativní, disjunktní

##### 4.5.1.3.6 typy alternací

#### 4.5.1.4 nerovnocennost alomorfů

### 4.3 MORFÉMIKA/MORFEMATIKA

= [část, disciplína] morfologie zabývající se lineární analýzou analýzou slova na morfémy {E: 273}

#### 4.3.1 Základní jednotky (nacionálně, třídění)

##### 4.3.1.1 morfém

###### 4.3.1.1.1 definice

- nejmenší, na sémantické úrovni dále nedělitelná jednotka jazyka mající povahu znaku ve smyslu jednoty označujícího a označovaného; relační jednotka určitelná na základě vztahu k sousedním jednotkám; je vyjadřován třídou alomorfů {E: 273}
- elementární znaky jazyka vydělitelné konfrontací různých kombinací, ve kterých se vyskytují (opakují) {K: 98}
- minimální jazykový znak, tj. minimální (abstraktní a systémová) významová a/nebo gramatické jednotka, chápaná, má-li alomorfy, jako třída alomorfů {Č: 122}
- nelineární jednotka, zobecněná, jejímiž představiteli jsou morfy; souhrn morfů téhož druhu totožných svým významem a formálně (foneticky) blízkých {M1: 180}
- typická interdisciplinární jednotka; nejmenší invariantní jazyková jednotka vydělitelná pomocí segmentačních pravidel, která je nositelkou významu a z něho vyplývajících funkcí {So: 9}
- segmenty s významem (...), a to koreňové, derivačné, modifikačné a gramatické {So: 29}
- invariantní[á], pomocí segmentačních pravidel vydělitelná[á] nejmenší[a] jazykov[á] jednotk[a], která je nositelkou významu a z něho vyplývajících funkcí {So: 30}
- 1) čistě vztahová jednotka 2) třída alomorfů 3) (minimální) spojení fonémů s vlastním významem {Er: 61}
- the second basic [the first = phoneme] unit in the expression system. (...) the unit (...) which enters into relationship with the content side; a group of one or more allomorphs which conform to certain, usually rather clearly definable, criteria of distribution and meaning {G: 11, 61}
- the smallest individually meaningful elements in the utterances of a language {H: 123}
- díky bilaterální povaze lze zkoumat jak formu (morf), tak význam (sémém, gramém) {E: 273}

###### 4.3.1.1.2 třídění

- třídění: na základě funkce: *lexikální* (pojmenovávají substance, děje, vlastnosti a okolnosti), *gramatické/relační* (pomocí nichž se vyjadřují gramatické významy), podle distribuce a povahy významu: *volné* a *vázané*, *autosémantické* a *synsémantické* (-les- a k), podle pozice ve slově (syntagmatická osa): *kořeny* a *afixy* (*prefixy*, *suffixy*, *postfixy* atd.) {E: 273}
- třídění: *vázané* a *volné* (do a nos), *synsémantické* (gramatické) a *autosémantické*: flektivní a nulové bývají synsémantické, též prázdné, ostatní autosémantické (s lex. významem); *flektivní* (≠ koncovka) × *slovotvorný*: a) kontinuální: nulový, supletivní × b) diskontinuitní {Č: 122}
- třídění: – *vázané* = nemohou se nikdy „vyskytnout v izolaci“; např. rozštěpený; – *separovaný* (mající jedinečnou vazbu); *reziduální* (-beere, -berry, -ina); – *diskrétní* (celek rozložitelný na fonémy), *kontinuální* (sekvence nefonematických segmentů určuje celý morfém, např. intonace, přízvuk); *reduplikované* (opakování celého morfému nebo jeho části), u opakování kořenu nebo kmene: někdy kompozita, podobné u onomatopoií, ovšem bez sémantických důsledků (fujfuj), české dělávávávat; *supletivní* (dobrý – lepší) {P: 111, 113}

##### 4.3.1.2 submorfém = konektém {K}

###### 4.3.1.2.1 definice

- diskrétna jazyková jednotka bez významu, ktorej funkciou je spájanie alebo rozširovanie morfému: a) interfixy, b) infixy {So: 47}
- invariantná jazyková jednotka vydeliteľná pomocou segmentačných pravidiel, ktorá nie je nositeľkou významu {So: 56}
- segmenty bez významu (...), ktoré fungujú pri spájaní morfému {So: 56}

#### 4.3.1.2.2 *trídění*

- *interfixy* (konektívne, interfigované) a *infixy* (tematické, rozširujúce, a to reziduálne a prehodnotené „degradované“ gramatické, lexikálne alebo derivačné (...)) {So: 29}

### 4.3.1.3 *morf*

#### 4.3.1.3.1 *definice*

- každá konkretizace morfému {K: 95}
- lineární jednotka; nejmenší část slovního tvaru členěného na bezprostřední složky mající svůj výraz a význam, vyčlenitelná na základě jejich opakování (v různých kombinacích) v jiných slovních tvarech {M1: 177}
- výrazová jednotka jazyka izolovaná při segmentaci mluveného n. psaného textu, řetězce fonémů a odpovídajícího grafického záznamu. Je vyjadřovacím prostředkem pro morfému. {E: 273}
- (grafická/fonická) diskretní (textová) realizace abstraktní jednotky morfému v textu (může být pro též morfému jedna nebo více. ...) Vztah morfému a morfu je tudíž též jako u typu a tokenu, resp. exempláře. {Č: 123}

#### 4.3.1.3.2 *třídění*

- třídění: podle funkce lexikální (kořeny – les), derivační (afixy – lesník) a gramatické (koncovky a pomocná slova – žen-a, smát se); podle povahy významu: *autosémantické*, *syntémantické*; podle pozice ve slově: *prefixální*, *kořenové*, *sufixální* (různého druhu nespa-l-a) {E: 273}

### 4.3.1.4 *submorf = konekt {K}*

#### 4.3.1.4.1 *definice*

- části, které při vyčlenění morfů ještě zbývají, mající pouze výraz, nikoli význam, elementy nižšího řádu; zpravidla se připojují k morfu předcházejícímu nebo následujícímu v tzv. *rozšířené variantu*; *konekty* ve funkci spojovat morfy a jejich komplexy ve slově {M1: 177}
- konkrétna realizácia submorfémy {So: 20}

#### 4.3.1.4.2 *třídění*

- *reziduální submorf* (dram:at-**urg**-Ø) {So: 31}

### 4.3.1.5 *alomorf = varianta morfému {M1}*

#### 4.3.1.5.1 *definice*

- různé konkretizace téhož morfému {K: 95}
- morf (rozložení jednotlivých morfů tvořících jediný morfému) podmíněný morfy sousedními: 1) formální strukturou sousedních morfů (fonémickým skladem začátku nebo zakončení, popř. slabičným skladem (sebrat × sbírat), 2) funkční stránkou těchto morfů (hoch = hoš před -i, -ích v NVL pl.) {M1: 181}
- fonologicky závislý variant tej istej morfémy (žiak-Ø/žiac-i) {So: 56}
- poziční varianta morfu; při segmentaci textu může mít jeden morfému několik realizací {E: 35}
- více textových realizací pro též morfému {E: 273}
- (distribuční/poziční) varianta morfému (vždy jedna z několika), pojatá jako systémová alternanta a součást třídy alomorfů {Č: 122}
- a variant of morpheme which occurs in certain definable environments {G: 61}
- (podle deskriptivistů:) 1) morf mající význam, 2) musí být v komplementární distribuci (...), pro češtinu by tak např. /-u ~ -i ~ -ovi ~ -e/ {K: 96}

- = *varianta morfému* – navzájem zaměnitelné v témž morfematickém okolí, závazná podmínka významové totožnosti a výrazové blízkosti {M1: 181}
- *úplná* (v okolí libovolných morfů, např. leda-, lec-) × *částečná* (jen v okolí jedněch morfů, v okolí druhých pouze jedna z částečných variant, např. -u/-ovi u živých bytostí, jinak pouze -u) {M1: 181}
- identitu znaku není s to narušit taková variabilita obsahu nebo výrazu, při níž se neztrácí souvislost mezi jednotlivými modifikacemi znaků {K: 97}
- porušení identity: *homonyma* (sémantická stránka), *synonyma* (výrazová stránka), vznik nového znaku (paralelně výrazová i významová): *město* × *místo* {K: 97}
- funkční variabilita morfů (alomorfie) spočívá ve funkční vzájemné substituci morfů, takové morfy jsou alomorfy (ps-át, piš-e, piš-0) téhož morfému {E: 273}
- distribuce alomorfů (variant morfémů) je řízena buď morfologicky (alomorfy prvního řádu, nebo fonologicky (alomorfy druhého řádu) {Er: 70}

#### **4.3.1.6 izomorfa**

##### **4.3.1.6.1 definice**

- morfologická a slovtvorná *izoglosa* (linie ohraničující výskyt určitého jevu), např. kravami – kravama – kravoma, slepice – slípka {E: 149}

#### **4.3.1.7 morfoném**

##### **4.3.1.7.1 definice**

- jako morfonologická jednotka myšlený úhrn fonémů (nebo skupin fonémů) účastnících se alternace, a to tak, že každý morfoném je reprezentován alternantou základní, výchozí, vzhledem k níž se druhá alternanta jeví jako odvozená; odtud směr alternace {M1: 182}
- obecná neznaková jednotka popisu, elementární zvukový segment vymezený na základě svého podílu na struktuře morfému; tvoří jej množina těch fonémů, které se mohou navzájem střídát na určitém místě morfému beze změny významu tohoto morfému; cílem popisu je ukázat prostřednictvím segmentálního zvukového složení morfémů vlastnosti a vztahy morfémové stavby daného jazyka {E: 165, 273}
- abstraktní třída fonémů se vzájemnou funkční substitucí uvnitř téhož morfému, resp. téže třídy; představuje tedy paradigma alternujících fonémů pro týž morfém {Č: 124}
- (podle Trubeckého:) soubor fonémů zúčastněných na určité alternaci {K: 96}

#### **4.3.2 Morfologická paradigmatika**

##### **4.3.2.1 slovní tvar = morfologická forma {E}**

- specifické syntagma, ve kterém gramatická složka, tj. morfém(y) determinuje (v sémantickém smyslu) složku lexikální – tj. morfologickou bázi {K: 97}
- = *morfologická forma*
- tvar slova (*utři*) nebo skupiny (*utřel jsem*) slov, ve kterém slova figurují v parole; podle morfologické formy se dělí: *ohebné*: skloňovatelné, časovatelné a *neohebné* {E: 137n}
- pravidelně se vyskytuje u všech slov určitého slovního druhu, popř. jejich většiny (absence sg. u pomnožných, analytického futura u dok.) {E: 137n}
- v češtině vysoká míra homonymie morfologické formy, gramatický význam až z kontextu {E: 137n}
- množina morfologických forem ≠ morfologické paradigma, to je tvořeno uspořádaným souborem morfologických forem nesoucích definovaný soubor gramatických významů {E: 137n}
- podle počtu slov tvořících morfologickou formu: *syntetická* – jednoslovné × *analytická* – složené tvary {E: 137n}

##### **4.3.2.1.1 složený/analytický/opisný/perifrastický tvar**

– taková forma flexivního paradigmatu, v níž jsou příslušné morfologické kategorie vyjádřeny kromě koncovky také (nebo dokonce výlučně) pomocí jednoho nebo více pomocných slov, v češtině v zásadě buď auxiliáry (tzv. pomocnými slovesy) (...), nebo partikulemi {E: 500}

– jejich komponenty mohou ve větě být od sebe odděleny v zásadě libovolným množstvím jiných slov; teprve tato výsledná kombinace jako celek vyjadřuje mluvnickou kategorii, která není samostatně přítomna v žádném z jednotlivých členů {E: 500}

#### **4.3.2.2 morfologická báze (slovního tvaru)**

– buď jeden kořenový morfém, nebo rozsáhlý komplex morfémů různého řádu vstupující do gramatiky hotová z lexikální roviny; derivovaná morfologická báze vzniká postupnou determinací (((uč-i-)tel-)sk-) {K: 98}

#### **4.3.2.3 slootovorný základ**

– odvozovací kmen, k němuž se připojuje slootovorný formant {E: 39}

#### **4.3.2.4 tvarotvorný základ**

– celé slovo kromě koncovky, u odvozených slov se obvykle kryje s odvozeným (tvarotvorným) kmenem {E: 39}

#### **4.3.2.5 (tvaroslovný/tvarotvorný) kmen**

– ta část slovního tvaru, k níž se připojují koncovky, a to *pádové* (pán-a, dobr-ého), *osobní* (nese-me), *rodové* (nesen-a), *infinitivní* (brá-t). Ze synchronního hlediska splývá pojem kmene u většiny substantiv až na výjimky (kuřat-a × kuřata) s pojmem *tvarotvorný základ*. {E: 215}

– kořenový morfém rozšířený o kmenotvorné morfémy; *odvozovací* (od něhož se tvoří dál) a *odvozený* {E: 39}

– u sloves sloves se *kmen* od tvarotvorného základu liší především u tvarů participií, u nichž je tzv. *sekundární kmen* tvořen od infinitivního kmene nekoncovým gramatickým afixem příslušného participia (bra-l-a, brá-n-a). {E: 215}

– *kmen* (vícemorfový základ) n. *báze* a *formant* (tj. morfematický jednoduchý nebo i složený afix) {Č: 149}

##### **4.3.2.5.1 jednoduchý × odvozený**

– *jednoduchý*, totožný s kořenem (žen-a, mil-ý, bý-t), n. *odvozený*, vzniklý pomocí kmenotvorných (odvozovacích) sufixů: učitel-k-a, měst-sk-ý, chod-íva-t {E: 215}

– kmen je nositelem hláskových alternací. {E: 215}

##### **4.3.2.5.2 infinitivní, přítomný, minulý**

– *infinitivní*: část neurčitého slovesného tvaru, ke které se připojuje koncovka infinitivu {E: 216}

– *přítomný*: část určitého slovesného tvaru, ke které se připojují osobní koncovky {E: 216}

– *minulý*: část neurčitého tvaru slovesného, ke které se připojují tvarotvorné přípony (a popř. koncovky rodové) l-ového participia (...); od kmene minulého se tvoří tvary participia l-ového a přechodníku minulého {E: 216}

##### **4.3.2.5.3 pasivní, imperativní**

– *pasivní* (proš-Ø-en), *imperativní* (tj. přítomný); Pasivní participium mívá mnohdy varianty kmenotvorné přípony, které se liší od variant kmene minulého. {E: 216}

#### **4.3.2.6 paradigma**

– 1) soubor jazykových jednotek propojený paradigmatickými vztahy, na něž se pohlíží jako na varianty jediné jednotky; 2) školsky a zúženě tabulka skloňování či časování, zvl. flexivních vzorů {E: 308}

= *paradigmatika*

– alespoň jeden člen se může proměnit, aniž se změní povaha vztahu {E: 551}

##### **4.3.2.6.1 kategoriální × formální**

– *kategoriální*: předepisující funkce, *formální*: nabízející inventář výrazů, které předepsané funkce obsluhují {E: 308}

##### **4.3.2.6.2 široké × úzké**

– *široké paradigma: úzká paradigmata* (kniha, časopis, noviny: kniha, -ou, ..., časopis, -em, ...) {Sn: 65n}

#### 4.3.2.6.3 morfologické paradigma (slovních druhů)

– systémy gramatických morfémů (a gramatických významů), které v slovních tvarech determinují morfologické báze tohoto slovního druhu {K: 98}

– soubor tvarů ohebného slova představující systém jeho gramatických kategorií {E: 308}

– své vlastní paradigma (subparadigma) ovšem může mít i slovní tvar; např. přechodníky, přičestí se dále modifikují morfémy s významem rodu a čísla {K: 98}

– z obecného m. p. lze vyčlenit paradigmata dílčí, např. subparadigmata sg. a pl., u sloves přítomní, imperativní ad. {E: 308}

#### 4.3.2.7 morfologické typy

– typické formy morfologických paradigmát závislé na obsahových i výrazových vlastnostech různých tříd morfologickýchází příslušného slovního druhu {K: 98}

– zahrnují (v rámci téhož slovního druhu) pravidla kombinatoriky morfémů a pravidla morfologických alternací {K: 98}

#### 4.3.2.8 slovtvorné typy

– typické formy derivovaných morfologickýchází vč. pravidel kombinatoriky morfémů a pravidel morfonologických alternací, navíc problematika onomaziologická a sémantická {K: 98}

#### 4.3.2.9 pád

– 1) pádový tvar, 2) pádový morf (pád koncovky), resp. pádový formant (deklinací) {E: 305}

#### 4.3.2.10 pádová polymorfie

– tzv. pádová polymorfie (synonymie) se v češtině využívá i u substantiv (u adj. nikoli); srov. dat. sg. -u/-i, -ovi u typu pán, muž, kde jde o distribuční využití v textu {E: 305}

#### 4.3.2.11 homomorfie

– (synchronně) soustavná *homonymie*, speciálně v paradigmatických vztazích (při existenci protikladu *formálního* a *kategoriálního* paradigmatu); lingvistika upřednostňuje pojmy *homomorfie*, popř. *polyfunkčnost* [před *synkretičností*] {E: 472 n}

### 4.3.3 Funkční a konstrukční typologie morfémů

#### 4.3.3.1 třídění

##### 4.3.3.1.1 vázané × nevázané / volné

– *vázané*: koreňové morfémy, které sa vždy viažu s derivačnými morfémami

× *nevázané / volné* = koreňové morfémy tvoriace lexému samostatne (proti) alebo tvoriace lexému s gramatickou morfémou (zim-a) {So: 15}

##### 4.3.3.1.2 samostatné

– *samostatné*: segmenty, ktoré sú v rámci lexémy alebo tvaru premiestniteľnými fonetickými slovami; (...) *sa, si* (...) *nie, sme, sa* {So: 15}

##### 4.3.3.1.3 funkční × prázdne

– mají určitý vztah k významu × nemají takový vztah {Sl7: 14n}

##### 4.3.3.1.4 jádrový

– jádro rozšířenéj morfémy (derivačná morfa *k* – mlad-un:*k*-ý) {So: 56}

##### 4.3.3.1.5 hypermorfém (rozšířený)

– vzniká rozšířením jádrovej morfémy o infix alebo o fonologický interfix; jádrová morféma rozšířená infixom (-un:*k*- mlad-un:*k*-ý) {So: 48, 56}

##### 4.3.3.1.6 bimorf

– (z hlediska syntaktického) předložka a pádový morf v předložkovém výrazu, který je relátorem {E: 305}

#### 4.3.3.2 kořen(y) / radix(y)

- nesamostatné morfémy vyjadřující elementární jazykové významy {K: 98}
- zřídka samostatné, většinou však nesamostatné morfy vyjadřující elementární (tj. nesložené) významy {M1: 178}
- nejmenšia diskrétna, bilaterálna jazyková jednotka, ktorá je nositeľkou lexikálneho významu a z neho vyplývajúcich funkcií {So: 31}
- nejjednodušší slovní základ, ďalej nečleniteľný, nesoucí významové jádro slova (...). K němu se připojují afixy (...). Často se však považuje kořen za vydělitelný nikoli na základě analýzy synchronní, nýbrž pojem formulovaný z hlediska praslovanského či indoevropského. {E: 236}
- kategorie stanoviteľná vesměs historicky, tedy z diachronního hlediska (...) a je společný nejen všem morfologickým formám slova (např. všem pádovým podobám), ale i slovům od něho odvozeným {E: 236}
- 4.3.3.2.1 reziduální, zbytkový**
  - jestliže můžeme ostatní části slova identifikovat z hlediska významového a výrazového jako morfémy afixální {R: 27}
- 4.3.3.2.2 jaderné (a tzv. rozšiřující elementy)**
  - případy kořenových morfémů, kdy sousedící s ním morfémy, buď prefixální, nebo sufixální, ztratily svou významovou funkci samostatnost a splývají s morfémem natolik, že jejich vzájemné vymezení je sporné; (...) „/s/věd/č/-i-t“, „do-/s/věd/č/-i-t“; *identifikované jaderné kořenové morfémy* (typu „-věd-“), které dovedeme identifikovat s odpovídajícími kořenovými morfémy v jiných slovech {R: 30}
- 4.3.3.2.3 reziduální**
  - (typu „-šet-“ ve slově „po-šet-i-l-ý“), u nichž význam pouze rekonstruujeme {R: 30}
- 4.3.3.3 afixy**
  - vždy jen nesamostatné morfy vyjadřující elementární nebo složené „gramatické“, tj. zobecněné významy {K: 98}
  - súhrnné označenie pre derivačné alebo flektívne morfémy rozličného druhu podľa pozície (spoločný názov pre *prefix, suffix, postfix*) {So: 56}
  - 4.3.3.3.1 gramatické/tvaroslovné/tvarotvorné**
    - 4.3.3.3.1.1 definice**
      - gramatické{K}/tvaroslovné (tvarotvorné){M1}
      - najmenšia diskrétna, bilaterálna a invariantná jazyková jednotka, ktorá je nositeľkou gramatického významu; jej funkciou je tvoriť nové gramatické tvary lexikálnych jednotiek v rámci gramatických kategórií a v základných tvaroch klasifikovať lexikálnu jednotku podľa slovných druhov {So: 45}
      - vyjadřují buď jediný gramatický význam (jediné séma), nebo celý komplex významů {M1: 178}
      - slouží ohýbání (flexi) slova, tj. skloňování a časování {M1: 178}
      - sú relačné, formové a kongruenčné; na rozdiel od koreňových morfém majú (...) relatívne uzavretý inventár morf a fungujú pri tvorení lexém a gramatických tvarov {So: 29}
      - prefixy, sufixy, postfixy {K: 98}
    - 4.3.3.3.1.2 třídění**
      - *podľa obsahu*: relačné, formové, kongruenčné, imperatívne, kondiciálové; *podľa pozície*: finálové, prefixálno-finálové (rozštiepené); variantné {So: 46}
    - 4.3.3.3.1.3 kumulativní morfémy**
      - celý komplex významů {K: 99} i {M1: 180}
      - gramatický význam (gramém), i když se kumuluje s jiným gramatickým významem v jediné koncovce, má možnost uplatňovat se v kontrastu s jinými tvary téhož slova i jiných slov téhož slovního druhu samostatně, jako jediný distinktivní významový rys na pozadí významových rysů společných (společného srovnávacího základu). {K: 99}

#### 4.3.3.3.2 slovotvorné

- bližšie určujú (determinujú) kořen {M1: 178}
- prefixy, sufixy, postfixy {K: 98}

#### 4.3.3.3.3 prefixy/předpony

- anteponovaná morfa pred koreňovou morfou (nie-č-o) {So: 56}
- majú miesto za kořenem, bezprostredne alebo za jiným sufixem determinujúcim kořen (a pozmenšujúcim jeho význam) {M1: 178}
- patrí ke slovotvorným formantům a je jedním ze slovotvorných morfémů vedle základových morfémů (kořen, kmen), sufixů a morfologické charakteristiky (...) {E: 341n}
- vzácnější je u českých prefixů funkce tvaroslovná: po-běží, po-nese; prefix a sufixy bývají označovány jako slovotvorný afix {E: 341n}
- je částí odvozeného slova, rozšiřuje se jím základové slovo tak, že předchází před kořenem a modifikuje význam základového slova (...), morfologická povaha základového slova (tj. jeho slovní druh a príslušnosť k paradigmatu) se prítom nemění {E: 341n}
- řada prefixů vznikla přehodnocením předložek; z hlediska formy podléhají prefixy proměnam: a) vokalizaci (...) b) dloužení vokálu; pokud se významová motivace základovým slovem necítí, pak se pouhá, byť vydělitelná forma za prefix nepovažuje: *nenávidět, závidět, podoba* {E: 341n}

#### 4.3.3.3.4 prefixoid

- téměř funkce prefixu {E: 505n}
- anteponovaná pôvodne koreňová morfa s derivačným významom, od prvých zložiek kompozit sa odlišuje oslabením lexikálneho významu a posilnením derivačného významu (tele-fón-Ø) {So: 56}

#### 4.3.3.3.5 sufixy/přípony

##### 4.3.3.3.5.1 finální/koncovka

- (pádové, osobní, infinitivní) stojí na konci slovního tvaru (následovat může postfix) = *koncovky* (pádové, osobní) {M1: 178}
- jeden ze slovotvorných formantů; společně s prefixem bývá označován jako slovotvorný afix; je zpravidla částí derivovaného (odvozeného) slova; následuje za odvozovacím kmenem a tvoří odvozený kmen (mil+ova-) nebo dovozené slovo (žák+yně); mění význam i morfologickou povahu základového slova {E: 468n}
- *finální sufixy* (koncovky), např. pádové, osobní koncovky; může stát před postfixem i radixem (pět-i-cípý) {K: 99}
- × *nefinální sufixy*, např. slovotvorné sufixy ohebných slov, gramatické morfémy transgresivů a participií (vedouc, vedouce, vedoucí) {K: 99}
- *koncovka*: u ohebných slov poslední tvarotvorný morfém, nebo u neohebných slovnědruhovú charakteristika (mlčk-y) {E: 39}
- *morfém tvarotvorný*: poslední morfém u tvarotvorných slov, neboli koncovka, v základním tvaru morfologická charakteristika (uč-i-tel-k-a), n. u neohebných slov slovnědruhovú charakteristika (mlč-k-y) {E: 39}
- *flexivní, finální*: vyjadřují buď jediný gramatický význam (jediné séma /po- v pojedú/), nebo celý komplex gramatických významů (dělník-ů)

##### 4.3.3.3.5.2 sufixy funkční × prázdné

- mající určitý vztah k významu („-i-“ v „modl-i-t-b(a)“) × nemající určitý vztah k významu („-t-“); *funkční*: průnik bezprostředních složek při postupu progresivním a regresivním {R: 32}

##### 4.3.3.3.5.3 tzv. složené sufixální morfémy

- jsou se základní podobou sufixálního morfému zcela shodné významově, ale liší se svým hláskovým složením tak, že jaderná část zvukového komplexu takového sufixálního morfému se nemění, ale rozšiřuje se vpředu (zleva) buď o *funkční*, nebo o prázdné *rozšiřující* elementy



{S17: 14n}

#### 4.3.3.3.5.3 *spolumorfém (sufixu)*

– morfológická charakteristika (žák + yň + e) {E: 468}

#### 4.3.3.3.5.4 *derivační*

– najmenšia diskretná, bilaterálna a invariantná jazyková jednotka, ktorá je nositeľkou slovotvorného, a to derivatvorného významu; jej funkciou je tvoriť nové lexikálne jednotky a systematizovať lexémy v rámci slovnej zásoby {So: 36}

#### 4.3.3.3.5.5 *modifikační*

– modifikujú lexikálny význam v rámci tvaru lexém; nositeľky významu gradácie, negácie a aspektu na úrovni lexikálno-gramatickej kategórie {So: 56}

#### 4.3.3.3.6 *suffixoid*

– pôvodne koreňová morfa s derivačným významom nachádzajúca sa po koreňovej morfe (osem-násť), prípadne pred gramatickou morfou (-oid – alkal-oid-Ø), od druhých zložiek kompozít sa odlišuje oslabením lexikálneho a posilnením derivačného významu {So: 56}

#### 4.3.3.3.7 *postfixy*

– pouze slovotvorné afíxy, pripojované až za gramatický sufix, teda k úplnému slovnímu tvaru (× prefixy také spojovány s celým slovním tvarem) {M1: 178}

– pôvodom slovanské alebo staročeské enklitické častice typu -k, -m(a), -n, -s(i), -to, -ť, -ž(e) a novější enklitika -hle a -pak, která v různé míře ztratila svůj enklitický charakter; slova s původními postfixy jsou dnes vnímána jako nemotivovaná (a – až, a – at', kdy – když ...) {E: 323n}

– postponovaná derivačná morfa, ktorá sa nepripája ku koreňovej morfe, ale nachádza sa po gramatickej morfe {So: 56}

#### 4.3.3.3.8 *postfixoid*

– pôvodne koreňová morfa s derivačným významom nachádzajúca sa po pôvodne gramatickej morfe (tret-i-krát) {So: 56}

#### 4.3.3.3.9 *infixy a interfixy*

##### 4.3.3.3.9.1 *infix*

– zvukový komplex dodatečne vkládaný do základu slova (vč. malilinký, dlouhanánský) {M1: 179}

– vkladná submorfa rozširujúca jadrovú morfu, ku ktorej sa pripája (infix un rozširuje jadrovú morfu k v adjektíve mlad-un:k-ý) {So: 56}

– rozširujúce morfémy, a to gramatické (t. j. tematické submorfémy) a derivačné morfémy / interfígované submorfémy) {So: 48}

– (vokalický, konsonantický), někdy za pomoci dalšího afíxu k utvoření slova, mluvnického tvaru (introflexe) {Č: 148}

##### 4.3.3.3.9.2 *interfix*

– segmenty vystupující mezi dvěma jednoduchými základy tvořícími složený základ (velk-o-město); nemají význam, tedy *submorfy*, resp. *konekty* {M1: 179}

– submorfa s funkciou spájať dve koreňové morfy v kompozitách (konektívny interfix, spájacia submorfa – vod-o-vod-Ø) alebo submorfa spájajúca dve morfy, ak sa jedna začína vokálom a druhá končí vokálom (fonologický interfix, interfígovaná submorfa j – Kóre:j-ec-Ø) {So: 56}

#### 4.3.3.3.10 *rarity afix*

– spájajúci sa s malým počtom koreňových morfém (smie-ch-Ø, ú-spe-ch-Ø) {So: 56}

#### 4.3.3.3.11 *unikátní afix*

– spájajúci sa s jednou koreňovou morfémou s neoslabeným motivačným vzťahom; má derivačný význam (kož-uch-Ø, ruk-áv-Ø) {So: 56}

#### 4.3.3.4 *afixoidy*

– spoločný názov pre *prefixoid*, *suffixoid*, *postfixoid* {So: 56}

– tvoří prechod mezi koreňovými a derivačnými morfémy, a to prefixoidmi a sufixoidmi {So: 29}

#### 4.3.3.5 lexikální morfémy

– (s Trubeckým:) morfematicky nerozložitelná slova nebo i slovní tvary: předložky (na, v, od), spojky (že, aš, jak, až), částice (ať, kéž, prý, už, ještě), některá cirkumstantiva (zde, dole, tam), interjekce (hle, hop, ach), zejm. osobní zájmena (já, ty, my, vy, mi, tě, nás), nesklonná přejatá subst. a adjektiva (prima, fit), pomocná slovesa (jsem, s) {K: 99n}

= *komplexní slovní tvar* – jediný morféma nositelem hodnoty lexikální i gramatické {K: 100}

#### 4.3.3.6 alternační morfémy

– alternace fungující jako jediný signál významového rozdílu mezi dvěma slovními tvary téže morfologické báze /Npl vojáci × Apl vojáci, Nsg ti × Dsg ti/ {K: 100}

#### 4.3.3.7 asociované (sdružené) morfémy = cirkumfixy {Č}, rozštěpené morfémy {Sl}

– (nerozložitelný) komplex významů vyjádřený zároveň dvěma morfémy (*po-ved-u*); oba se podílejí na vyjadřování příslušného významu; tvoří společně jedinou bezprostřední složku {K: 100}

= *rozštěpený morféma* / *cirkumfix*

– není nutné zavádět (...), jde zde prostě o dva funkčně spjaté morfémy, o prefixálně sufixální tvoření, které probíhá současně {Sl2: 95}

= *cirkumfix* (diskontinuitní afix) {Č: 149}

#### 4.3.3.8 konektémy (konekt – alokonekt) = submorfémy {M1, So}

– tzv. *spojovací morfémy*

– 1) mohou mít konkretizace (*ved-Ø-u* × *ved-e-š*), 2) mohou být sekundárně nositeli významu (meziparadigmaticky: *sed-ě-t* × *sed-a-t*) {K: 101n}

– příznačná pro slovesa, adjektiva, zájmena, v kompozitech, některých plurálových tvarech a okrajových typech substantivní flexe (*žen-á-m*, *kost-e-m*, *kry-j-e-š*, *kup-o-v-a-t*, *uděl-a-v*, *děl-aj-í*, *sáz-ej-í* ...) {K: 101n}

– sekundární konektivní funkce slovotvorných morfémy (*sokrat-ov-ský* × *platón-ský*) {K: 101n}

##### 4.3.3.8.1 redundantní, prázdné morfémy

– tvoří s následujícím slovotvorným sufixem jedinou bezprostřední složku v rámci morfologické báze (((kav-)ár-n-)a, ((zn-a)-tel-n-)ý) {K: 102}

##### 4.3.3.8.2 složený slovotvorný morféma

– kdy bezprostřední složkou morfologické báze je sekvence slovotvorných morfémy, v níž před tzv. *jaderným morfémem* (posledním v sekvenci) předchází slovotvorný morféma se sekundární konektivní funkcí {K: 103}

##### 4.3.3.8.3 kmenotvorný morféma

#### 4.3.3.9 kmenový konektém / kmenotvorný morféma {E}

a) konektém tvoří těsnější celek s předcházející morfologickouází

= *kmenový konektém* / kmenotvorná přípona {M1}

– má proporcionální výskyt (vč. alternací); verbální, adjektivní a zájmenná flexe, okrajově substantivní; je bezprostřední složkou kmene {K: 103}

b) konektém tvoří těsnější celek s následujícím gramatickým sufixem

= *složený gramatický sufix*

– sám bezprostřední složkou kmene není: *žen(-á-m)*, *chlap(-e-ch)* {K: 103}

– rozšíření kořene na kmen (jednoduchý, nebo víceúrovňový) {E: 39}

#### 4.3.4 Identifikace alomorfů

– *morf* = konkretizační výrazová forma morfémy, *konekt* = konkretizační forma konektému {K: 104}

##### 4.3.4.1 alomorf (alokonekt) téhož morfémy (podmínky):

- 1) identita významu nebo funkce (u konektémů), 2) materiální souvislost výrazu, 3) nejčastěji komplementární distribuce {K: 104}
- 1) mají-li nějaký společný význam, 2) jsou-li ve vztahu komplementární distribuce, podmíněné nějakým fonologickým nebo morfologickým faktorem; *komplementární distribuce* (dvou elementů) jestliže se jeden z nich vyskytuje v takovém okolí, v jakém se nikdy nevyskytuje druhý, tj. jestliže neexistuje okolí, v kterém by se mohly vyskytovat oba {SI2: 96}
- Two elements can be considered as allomorphs of the same morpheme if: (1) they have a common meaning, (2) they are in complementary distribution, and (3) they occur in parallel formations. {G: 88}
- významový rozdíl identického/homonymního sufixu (let-ec × slep-ec) je dán slootovornou bází, v obou případech „nositel příznaku“ a) dynamického (letec), b) statického (slepec) {K: 104}
- u morfémů kořených je jednomorfémovost jev periferní (m-í-t, d-á-t) {K: 105}
- nepovažujeme za alomorfy různé jednofonémové morfy (především vokalické) paralelních morfologických typů {K: 106}
- výraz jednoho z alomorfů (alokonektů) může být buď zčásti, nebo zcela nulový (transgr. /-vš- ~ -v- ~ -š- ~ Ø/) {K: 106}

#### 4.3.4.2 typy alomorfů

##### 4.3.4.2.1 nekomplementární

- pánův/pánů, inf. -ti/-t, (níž-Ø a níž-e *stylistické*), *různomorfémové* – bez materiální souvislosti {K: 107}

##### 4.3.4.2.2 komplementární

###### 4.3.4.2.2.1 paralelní

- komplementární (i neúplně) v rámci téhož gramatického významu (Asg /-Ø, -a, -u, -i, -o, -e, í/) {K: 107}
- buď v různých morfologických typech téhož paradigmatu, nebo v komplementárních tvarech téhož morfologického typu (/e, -u/ pane, vojáku) {K: 107}

###### 4.3.4.2.2.2 synonymní

komplementární jen v poměru k různým typům, v rámci téhož morfologického typu jsou volitelné (jazyk-u/jazyc-e, píš-u/píš-i) {K: 107}

###### 4.3.4.2.2.3 supletivní

- komplementárnost se projevuje v morfologické bázi, souboru tvarů téhož morfologického typu nebo v gramatické části téhož subparadigmatu slovního tvaru {K: 107}
- především kořenné morfémy (js- / bud- / bi-; já / mne / mně; -a, -ouc; -i- / -e- v slyším, slyšet) {K: 107}
- morfy s totožným významem, ale zcela rozdílným výrazem patřící k různým morfémům (ber-, br-); výjimečně totožný význam a jedn realizovaný fonémovou nulou (nes-Ø – děla-v); homonymní (nos-it × nos-Ø) {M1: 190}

### 4.3.5 Souvztažné pojmy

#### 4.3.5.1 nulový morfém (konektém) / morfémová nula {E}

- jestliže má vedle sebe aspoň jeden paralelní nebo supletivní morfém (konektém), jehož výraz má fonémickou strukturu; proto vždy paralelní nebo supletivní morfém (konektém) (pán-Ø, žen-Ø, nes-Ø-ti, nes-e-Ø, pros-í-Ø, nes-Ø-l-Ø + Ø) {K: 108}
- nulový morfém je třeba lišit od nulového alomorfu (alokonektu) – např. v tvarech /hůř-e/ – /hůř-Ø/, /ved-e-š/ – /ved-Ø-ou/ {K: 108}
- = *morf nulový (morfémová nula)*
- abstrakce na pozadí celého morfologického paradigmatu, má distinktivní funkci (pán-Ø × -a; nes-Ø-Ø × tisk-n-i) {E: 273}

– nulovým morfémem rozumějí představitelé všech směrů relevantní nepřítomnost jakéhokoli gramatického morfému {Sn: 203}

#### **4.3.5.2 morfémový/morfematický šev**

– nachází se na místě styku (hranici) dvou morfémů; může tak vzniknout buď hlásková alternace (...), nebo stojí-li vedle sebe dva stejné nebo blízké fonémy, může docházet k *fúzi*, při níž je některý z fonémů tvořících příslušné morfy redukován, a to splnutím hlásek (rus > rus+ský > ruský), zjednodušením souhláskových skupin (češ-ský > český, ob-vléci > obléci), nebo v kombinaci s hláskovými alternacemi překrýváním hlásek slovotvorného základu a slovotvorného formantu (kameník+ský > kamenický) {E: 468, 475}

#### **4.3.5.3 morfematický / morfémový uzel**

– překrývání a splývání morfémových švíků {So: 28}

– při styku dvou morfémů, z nichž jeden končí a druhý začíná stejnou (nebo foneticky blízkou) hláskou (...) jedna hláska je součástí obou morfémů {R: 31}

– vzniká na morfémovém švu, teoreticky pociťuje mluvčí existenci původního morfému, který ovšem nemá příslušnou realizaci {E: 475}

#### **4.3.5.4 perintegrace / přerozdělení {E}**

– změna morfematické segmentace slova; dochází vlivem oslabení původní slovotvorné motivace a s ní související segmentace k sekundárnímu rozložení morfémů ve slově {E: 313}

– zmena, zánik alebo prehodenie pôvodných morfémových hraníc (... krídl-o, dar-Ø, daň-Ø ...) {So: 24}

– prefixální morfém považujeme za perintegrovanou součást kořenového morfému v případech, kde stupeň lexikalizace je takový, že etymologicky vydělitelný kořenový morfém již v současné slovní zásobě nefunguje jako samostatná lexikální jednotka (zdvih-a-t, rozpaky); jaký[sí] spojovac[í] můst[ek] mezi identifikací významovou a výrazovou, protože prolíná obě tyto identifikační roviny {S15: 362}

#### **4.3.5.5 deetymologizace**

– globalizace struktury slova vedoucí až k úplné ztrátě členitelnosti; příčiny: např. oddálení významu, zejména přenesením, modifikace významu, nejčastěji prefixací, hláskové změny neproduktivní a málo frekventované aj. {S15: 362}

#### **4.3.5.6 \*trunkace/trunkácia**

– odseknutie alebo vynechanie morfy, resp. jej časti (meteor-o-log-ic:k-ý > mete-o-serv, nos-o-rož-ec-Ø, nos-o-rož-í) {So: 56}

#### **4.3.5.7 distribuce (libovolného morfému)**

– souhrn všech typů morfémů, v jejichž okolí se může daný morfém vyskytovat {S12: 95}

#### **4.3.5.8 centrum a periferie**

##### **4.3.5.8.1 centrum**

– oblast elementů maximálně zapojených do jazykového systému všemi charakteristickými rysy; z hlediska syntagmatické osy (...) centrální (...): jednoduché, jednoznačně vydělitelné kořenové morfémy, tj. které nejsou složeny z jaderného morfému a rozšiřujících elementů; míra perifernosti kořenového morfému bude vzrůstat s počtem jeho konstituentů a jejich kombinacemi {S17: 12–14}

##### **4.3.5.8.2 periferie**

– oblast elementů ne plně, méně, nebo jen sporadicky využitých v jazykovém systému {S17: 12–14}

#### **4.3.5.9 morfostylém**

= např. využití hovorové tvaroslovné dublety ve vědeckém textu {E: 463}

#### **4.3.5.10 odpovědnost a kompetence morfému**

##### **4.3.5.10.1 odpovědnost morfému**

– podíl složek (morfémů) na významu konstrukce (tvaru); Je-li monomorfematický celý slovesný tvar (nes), pak jediný morfém tvaru, morfém kmenový, je odpověden nejen za lexikální význam L, ale i za celý význam gramatický G. {Ro: 53n, 66}

#### 4.3.5.10.1 kompetence morfému

– významová potence každého morfému (kontextově nepodmíněná významová konstanta) {Ro: 67}

#### 4.3.5.11 fúze

– U složitějších morfematických útvarů (kmenů, slovních forem) vedly syntagmatické hláskové změny nezřídka k tzv. fúzi, tj. ke splývání morfů – zejména k zániku morfematických švů mezi kořenem (R) a formantem (derivačním sufixem) (R + F) a mezi formantem a gramatickým sufixem (F + Ag).; Mnohá česká slova, jež se při synchronické analýze jeví jako jednoduchá (člověk, nevěsta, medvěd aj.) jsou od původu složeniny. {Er: 74}

### 4.4 MORFOTAKTIKA

#### 4.4.1 definice

– část morfologie sledující kombinatoriku, lineární uspořádání a distribuci morfémů především v rámci slovtvorby {E: 273}

– studium systematického uspořádání morfémů v parole {Č: 125}

– Tactics refers to the ordering of elements at any level of linguistic structure, and the rules that govern this ordering. Morphotactics, then, deals with the possible sequencing of morphemes within words. {Ha: 335}

#### 4.3.2 obecné poznámky

– z hlediska fonologické formy (morfy) jednofonémové × vícefónémové {M1: 179}

– pozice fonémů v morféměch se někdy dělí na *silné* (rozlišující význam dvou různých morfémů) a *slabé* (ostatní) {P: 116}

– u flektivních jazyků poměrně malý počet morfů vázaných na nějakou paradigmatickou třídu (kumulace funkcí a koncovek); častá homonymie a synonymie jednotlivých koncovek {P: 150}

– v žádném jazyce se neshoduje počet slov a morfémů. Počet morfémů ve vyspělých jazycích čítá řádově několik tisíc, počet slov dosahuje několika desítek, či dokonce set tisíc.“ {Sn: 205}

– kombinace plnovýznamových morfémů uvnitř slova jsou do jisté míry analogické kombinacím slov ve slovních spojeních nebo ve větách.“ {Sn: 212}

#### 4.3.3 fonematická struktury morfů

##### 4.3.3.1 čeština

– fonémová struktura morfů je v zásadě dána jejich funkčním druhem:

*kořenové* – zřídka jednofonémové, častěji dvoufonémové a zvláště třífonémové

*sufixální morfy* – slovtvorné zpravidla jednofonémové (souhláskové) nebo dvoufonémové, řidčeji třífonémové a vícefónémové

*koncovkové* – pro subst. jednofonémové (samohláskové a dvojhláskové), dvoufonémové (typu V +K, výjimečně K + V) a trojfonémové (typu V + K +V); pro slovesa jednofonémové (souhláskové a samohláskové); realizované nulou / nulové

*prefixální morfy* – jednofonémové (samohláskové, souhláskové), dvoufonémové (V +K a zvl. K + V, ojediněle K + K), řidčeji trojfonémové (K + V +K, K + K + V) a vícefónémové {M1: 179}

– počet morfémů v českém slově je omezen, slova mající více než 9 morfémů jsou nepravděpodobná {P: 126}

#### **4.3.3.2 slovenština**

– gramatický tvar sa skládá z 1–n K[oreňová morfa], 0–n D[erivačná], 0–n M[odifikačná], 0–n G[ramatická] {So: 51}

### **4.5 MORFONOLOGIE**

#### **4.5.1 Vývoj a obsah morfonologie**

##### **4.5.1.1 definice**

– sleduje fonémické složení morfů {E: 273}

– morfo(fo)nologie (vznik, povaha a distribuce morfonémů) {Č: 124}

– označení prvků teoretických základních forem morfémů {K: 112}

– might include ‘changes’ in the form of morphemes that may be described by phonetically natural or unnatural rules, by rules that effect fusion between adjacent phonemes or do not, by rules that are subject to grammatical or lexical limitations on their applicability, and by rules describing correspondences that are phonologically unnatural or mere historical residues (Carstairs-McCarthy, 1983: 237–238) {Co: 318}

– has as its basis the comparison of (1) the shapes that words/morphemes have in different grammatical and lexical contexts and (2) the shapes that morphemes have in different but related lexical items {Co: 318n}

##### **4.5.1.2 morfoném**

– (podle Trubeckého) třída fonémů podílejících se na určité alternaci v morfému, a tedy na celkové stavbě morfému; reprezentován příslušnými fonémy; dva nebo více alternujících fonémů {K: 108}

– fonémickou stavbu, tj. fonémické reprezentace určitého morfonému, mají až konkretizace morfémů, tj. morfy a alomorfy {K: 108}

– (podle Komárka) alternační třídy jednoprvkové i víceprvkové vymežitelné vztahem vždy k určitému morfému {K: 108}

##### **4.5.1.3 morfofoném**

– (podle Bloomfielda a jeho následovatelů) hypotetické fónické prvky (konstrukty), sloužící jako stavební materiál teoretických základních forem. Jejich systém má za základ systém fonémů s jejich protiklady, ale není s ním plně identický, nýbrž v detailech se od něho liší {K: 109}

– (podle Harrise) morfofonémický symbol reprezentující třídu fonémů a definovaný soupisem fonémů, které se v témž morfému na témž místě vyskytují v různém okolí {K: 110}

– jako distribuce distinktivních rysů (Halle ad.), kdy [existuje] přímý vztah mezi morfonologickou a fonetickou reprezentací morfému {K: 111}

– skutečnou elementární jednotkou zvukového plánu jazyka nejsou fonologické distinktivní rysy, nýbrž fonémy jakožto minimální segmenty. Distinktivní rysy jsou jen vlastnosti fonémů, různou měrou realizovatelné a rozlišitelné v komplexní struktuře jejich fónických konkretizací. {K: 112}

#### **4.5.1 Hláskové a morfologické alternace**

##### **4.5.1.1 hláskové alternace**

– pravidelné střídání vybraných fonémů, popř. skupin fonémů, na němž je založena alomorfie; ryze *morfologické* (filolog – filozof), které se vyskytují pouze v rámci systému flexe, čistě *slovotvorné* (polévka – polívka, k nimž dochází výhradně při tvoření slov), a vyskytující se *jak ve flexi, tak ve slovotvorbě* (jméno – jmen, kámen – kamínek) {E: 35n}

- klasifikace alternujících hlásek: *alternace vokální* a) kvantitativní, b) kvalitativní (vejce – vajec), c) kvantitativně – kvalitativní (třást – třese) a d) zánikové (pes – pØsa), *konsonantické* a) párové (měkčení /slon – sloň/, tvrdnutí (zeď – zdem)), b) nepárové (matka – matce), c) zánikové (hýbat – hØnout) a d) skupinové (český – čeští, rozjezdít – rozježděn), *smíšené* (stát – stojí, vát – věje) {E: 35n}

- mohou se vyskytovat ve všech typech morfémů {E: 35n}

#### 4.5.1.2 morfonologické alternace / střídání fonémů

- fonologicky nepodmíněná zákonitá záměna téhož morfému při tvoření slov a tvarů; výskyt toho či onoho alomorfu nemůže být objasněn fonologickými zákony současného jazyka, nýbrž pouze historicky. Z hlediska současného jazyka je podmíněn pouze sousedními morfy v struktuře slovního tvaru. {M1: 182}

- alternanty tvoří v jazykovém povědomí jeden *morfoném*, tj. jako morfonologická jednotka myšlený úhrn fonémů (nebo skupin fonémů) účastnících se alternace, a to tak, že každý morfoném je reprezentován alternantou základní, výchozí, vzhledem k níž se druhá alternanta jeví jako odvozená; odtud směr alternace {M1: 182}

#### 4.5.1.3 alternační řada

##### 4.5.1.3.1 uspořádanost × neuspořádanost

- fonémy účastnící se alternace; zpravidla dvojčlenná, řidčeji vícečlenná:

- 1) *neuspořádaná dvojice* = n-tice alternant, tzn. bez určitého pořadí, tedy t/t' nebo t'/t, k/c/č/ nebo k/č/c (...)

- 2) *uspořádaná dvojice* = n-tice alternant charakterizovaná určitým směrem alternace, s pevným pořadím alternant (může být odrazem reálných nebo domnělých hláskových procesů, jejich projekcí do synchronie {M1: 182}

##### 4.5.1.3.2 základní varianta

- morfonologicky podle funkce při stavbě slova nebo tvaru: základní ta, která vystupuje v základovém (fundujícím) slově slovotvorné (derivační) dvojice (hlas – hlásek = a/á, ruka – ruční = k/č) {M1: 182}

##### 4.5.1.3.3 princip většinové podoby

- převládající, dominantní: nést / nesl, nesu ... > e/é {M1: 184}

##### 4.5.1.3.4 fonémová nula / nulový morf

- člen alternační řady i spojení dvou fonémů (st/šť), jedním z členů i fonémová nula (nulový morf) {M1: 184}

##### 4.5.1.3.5 alternativy korelativní, disjunktní

- alternativy korelativní (přítomnost fonologického rysu × nepřítomnost) – disjunktní (jeden fonologický rys × jiný) {M1: 184}

##### 4.5.1.3.6 typy alternací

- A) samohláskové, B) souhláskové, C) souhláskových skupin, D) smíšených skupin {M1: 184–189}

#### 4.5.1.4 nerovnocennost alomorfů

- a) podle stupně využití (vlc ve 2 tvarech × vlk v 11)

- b) stupně kontextové motivace (vlc silnější než vlk)

- c) z hlediska prediktability (zvon-ek > zvon-c-ích × zvon-c-ích > zvon-ec, zvon-ek {K: 113}

- *plná* a *redukována* varianta (konektém -aj- – -a-) {K: 113}

- třída morfonémů „tvrdých“, „měkkých“, „obojetných“ (průnik předchozích) {K: 106}

## 4.6 Shrnutí

Kapitola nejprve velmi stručně uvádí do problematiky tzv. stratifikační lingvistiky, z níž připomíná některá běžnější kritéria pro vymezení jednotlivých jazykových rovin, a představuje z perspektivy začlenění jednotky morfému vybrané (užívané i pouze navrhované)

hierarchické modely. Pro navazující přehled poznatků ze sféry *formální morfologie*, *morfematiky*, *morfotaktiky* a *morfonologie* je morfém chápán v souladu s nejběžnějším pojetím, tedy jako základní jednotka roviny morfologické, resp. morfematické. Repetitorium především poskytuje teoretické zázemí pro návrh formálního počítačového zpracování morfematiky, kromě toho dává vyniknout pluralitě terminologie i vlastních pojetí, tak jak ji zprostředková exerpce z klíčových děl zhruba poslední čtyřiceti let. V relativní úplnosti je zachycena především teoretická koncepce M. Komárka a výklady z Encyklopedického slovníku češtiny.



## 5 Dosavadní empirická (zejm. lexikografická) zpracování morfematiky

### 5.1 Nástin vývojového driftu

Empirickým výzkumem morfematické stavby konkrétních jazyků se pravděpodobně jako první zabývali američtí deskriptivisté (srov. BOSÁK – BUZÁSSYOVÁ 1985: 30). Jejich motivací bylo vytvořit první popisy domorodých jazyků, které neměly své písmo. Předmětem analýz byly většinou přepisy konkrétních promluv, cílem takového snažení vymezit funkční jednotky zkoumaného jazyka a popsat jejich distribuci (odtud označení amerických strukturalistů jako *distribucionalistů*). Postupem času vznikla podrobná metoda segmentace, která byla dostatečně obecná, tak aby vyhovovala všem typologickým rysům popisovaných jazyků (viz např. GLEASON 1969). Pro jazyky s dostatečnou písemnou tradicí se při ověřování teoretických postulátů morfematické analýzy nejprve využívala cílená excerpce z odborných<sup>56</sup> (např. SLAVÍČKOVÁ 1962), beletristických (např. SABOL 1973), ale i žánrově rozmanitých textů (SLAVÍČKOVÁ 1962, SABOL 1979).

Přirozená potřeba zvýšit kredibilitu publikovaných výsledků přiváděla badatele k úvahám o reprezentativnosti výchozích jazykových dat. Přechodným řešením byla snaha využít údaje z frekvenčních slovníků (SLAVÍČKOVÁ 1965, 1967). Jejich existence pro některé jazyky nebyla – na rozdíl od jiných typů slovníků – vůbec samozřejmá, kromě toho mnohdy nedostačovaly svým rozsahem. Morfematické slovníky ze šedesátých a sedmdesátých let 20. století (WORTH – KOZAK – JOHNSON 1970, RMSČ/SLAVÍČKOVÁ 1975, AUGST 1975) tak byly založeny na dostatečně rozsáhlém materiálu (minimálně 50–60 tisíc lexémů) čerpaném z různých lexikálních zdrojů. Pro analýzu systému jazyka (centra) to bylo optimální množství, z hlediska ruční analýzy naopak horní, ještě přijatelná mez. V některých případech byla proto poprvé využita počítačová technika. Většinou šlo o pomoc s tříděním a přípravou pro tisk, někdy však stroje pomáhaly i s prvotní analýzou (WORTH – KOZAK – JOHNSON 1970).

Zcela výjimečně<sup>57</sup> (BERGENHOLTZ 1976) se morfematická problematika začala zkoumat i na autentickém korpusovém materiálu, který musel být z podstaty věci uložen na datových nosičích a obsluhován strojově. Počátky české korpusové lingvistiky spjaté s tzv. *Korpusem věcného stylu* (viz KRÁLÍK 2009) o velikosti 540 tisíc slov (tokenů), který sestavil tým Marie Těšitelové v Ústavu pro jazyk český během první poloviny sedmdesátých let 20. století, s výzkumem morfematiky nijak nesouvisely. Monografie *Kvantitativní charakteristiky současné češtiny* (TĚŠITELOVÁ ET AL. 1985), jež představuje shrnutí všech souvisejících výzkumných aktivit, neobsahuje o morfěmech ani zmínku<sup>58</sup>. První údaje o morfematických aspektech češtiny založené na korpusových datech pocházejí teprve z loňského roku (STATČ). Týkají se absolutní frekvence nejčastějších morfémů získaných ruční analýzou<sup>59</sup> vzorků o velikosti jednoho tisíce nejfrekventovanějších substantiv, adjektiv, verb a adverbii. Data pocházela ze stomilionového korpusu SYN2005, pokrytí textu (součet frekvencí) by mělo být zhruba třicetiprocentní.

Jakékoli bližší informace o použité metodě segmentace nejsou známy (v tzv. bezhodnotové lingvistice (viz kap. 0.2) je to součást záměru). Navzdory ujištění, že byla analýza provedena ručně, jsou vyznačené hranice morfémů značně „podspeifikovány“, jinými slovy stanoveny příliš vágně. U substantiv jsou za nedělitelné segmenty pokládány např. *vztah* (× vz-tah), *důvod* (× dů-vod), *informac* (× inform-ac), *ínek* (ín-ek) nebo *pravd* (prav-d). Mezi adjektivy jsou za samostatné morfémy považovány jednotky jako *dalš* (× dal-š), *možn* (mož-n), *hranič* (× hran-ič) či *základ* (× zá-klad), údajnými kořeny jsou dokonce

<sup>56</sup> Slavičková (1962) prováděla první morfematické analýzy na excerptech ruštiny odborného stylu.

<sup>57</sup> Pro exploraci sémantické stránky morfémů v němčině byl využita reprezentativní část korpusu o velikosti 300 tisíc textových slov nazvaná LIMAS-Korpus, pocházející z celkového množství asi tří milionů slov.

<sup>58</sup> Srov. pasáže o E. Slavičkové v kap. 0.1.

<sup>59</sup> O prvních pokusech o automatickou (neřízenou) analýzu morfémů v korpusech češtiny viz 6.2.

třimorfémové shluky: *důležit* (× dů-lež-it) atd. U zbývajících seznamů pro slovesa a adverbia je to velmi podobné. Pokud je toto výsledek nejnovějších korpusových metod, potom je pozice RMSČ jako etalonu morfematického popisu češtiny prakticky neotřesitelná, a to i dlouhých 35 let od vydání. Nejen z tohoto důvodu má smysl se ke starším morfematickým (a derivačním) slovníkům vracet a uvědomit si jejich přínos i trvalý význam.

Jak připomenul před časem J. Marvan, „[d]odnes ale, na rozdíl od jiných Slovanů, derivační slovník nemáme. Všechny naše staleté primáty tak byly bohužel proplytvány 19. a zejména 20. stoletím“ (MARVAN 2006: 146). S drobnou výhradou, kterou jsem uvedl v kap. 3.5, nezbyvá než dát Marvanovi za pravdu a také tyto jinoslovanské výzvy adresované lingvistické bohemistice zahrnout do následujícího stručného přehledu doplněného vždy faksimilní ukázkou několika hesel, resp. hnízd. Jednotlivá díla jsou představována podle jazyků a následně chronologicky.

## 5.2 Přehled vybraných lexikografických přístupů k morfematice

### 5.2.1 Čeština

#### 5.2.1.1 Slavičková: *Retrográdní morfematický slovník češtiny (1975)*

V přehledové stati o vývoji české gramatografie stanovuje Petr Karlík význam klíčového díla české morfematiky takto: „*Před Komárkovými Příspěvky vychází taky slovníkové zpracování morfematiky češtiny od Eleonory Slavičkové (...)*“ (PLESKALOVÁ ET AL. 2007: 98n) V jiné části téhož díla, kapitole věnované proměnám české lexikografie od počátků až po současnost, je o *Retrográdnímu morfematickému slovníku češtiny* zmínka v souvislosti s jeho formálním zpracováním. Morfematicky rozčleněná slovní zásoba je seřazena způsobem a tergo, tedy podle zakončení. Internetové vyhledávače na dotaz obsahující název díla vracejí jen několik málo desítek odkazů. Žádný neobsahuje podrobnější zmínky, většinou jde jen o položku v bibliografii nebo záznam z katalogu knihovny. Odborná i laická recepce je tedy shodně zanedbatelná. K většímu věhlasu práce nepomohla ani vstupní, zhruba třicetistránková *Prolegomena* v češtině, ruštině, angličtině a francouzštině, popisující mikrostrukturu a makrostrukturu slovníku a zahrnující i *Komentář k teoretické koncepci morfémové analýzy a seznamům morfémů*<sup>60</sup>. Promítnutí konkrétních principů do jednotlivých segmentací ovšem není neproblematické. Jak ve své dobové recenzi správně poznamenal D. Šlosar, „[n]ěkdy si ovšem uživatel není jist, zda nejde o tiskovou chybu, srovnej např. rozdělení tř-i-válc-ov-ý × dv-oj-vál-c-ov-ý“ (ŠLOSAR 1977).

Dílčí nekonzistence ve zpracování zřejmě zavinila předtisková příprava, protože sazba zřetelně připomíná strojopis<sup>61</sup>, i když byl slovník připraven za pomoci počítači ZPA-600. Tisk v nákladu 700 kusů provedl Vojenský zeměpisný ústav a technická omezení umožnila publikovat jen část zpracovaných kombinačních sestav. Konečná knižní podoba tak obsahuje 1) retrográdní slovník morfematicky členěných slov opatřený jedno- až dvoumístných indexem kódujícím stylistické a sémantické charakteristiky, 2) inventární slovníky českých morfémů. Do nich byly zařazeny:

- A Seznam kořenových morfémů, vyčleněných ze současné české stylisticky neutrální slovní zásoby (se všemi variantami)
- B Seznam nejfrekventovanějších kořenových morfémů (s četností nad 50 výskytů)
- C Celkový seznam sufixálních morfémů (jaderných se všemi variantami)
- D Seznam jaderných sufixálních morfémů substantivních, verbálních, adjektivních a adverbálních
- E Seznam jaderných sufixálních morfémů ostatních slovních druhů (numeralia, pronomina, partikule) sysémantické povahy

<sup>60</sup> Popsané zásady morfematické analýzy zpřesňují koncepci publikovanou již dříve (SLAVÍČKOVÁ 1962 a 1967) a doplňují ji příkladovým materiálem čerpaným z hesláře. Podstatné rysy zásad uvádím a kriticky hodnotím v 6.1.

<sup>61</sup> Indicií je např. občasné ruční zalomení delšího řádku.

- F Celkové frekvenční pořadí nejfrekventovanějších jaderných sufixálních morfémů
- G Frekvenční pořadí nejfrekventovanějších jaderných sufixálních morfémů podle slovních druhů
- H Seznam prefixálních morfémů a) domácích, b) cizích, c) expresivních
- I Frekvenční pořadí nejfrekventovanějších domácích prefixálních morfémů
- J Seznam dvojic domácích prefixálních morfémů
- K Seznam trojic domácích prefixálních morfémů
- L Seznam homografních kořenových morfémů domácích (neutrálních, expresivních a archaických) i cizích

Heslář čítající 63 634 lexikálních jednotek pochází především z *Česko-německého slovníku* J. Volného, doplňujícími prameny byly však také *Česko-německý slovník* H. Siebenscheina, *Stručný etymologický slovník* autorů Holuba a Kopečného (jako zdroj oborové terminologie a cizích slov) a rovněž *Slovník spisovného jazyka českého* (k doplnění některých novějších slov a naopak jako pomůcka pro vypuštění výrazů archaických, slangových, nářečních, vypjatě expresivních a úzce odborných). Problematickým rysem výběru (zejm. z hlediska kvantitativní analýzy uvedené v kap. 7) se stalo rozhodnutí zahrnout i jiné než základní podoby (prototypické slovníkové tvary, lemmata) a naopak (maximální délkou řádku) vynucené nerozlišování reflexivity u sloves (morfémem se, si). Týká se to např. výběrového zařazení dvou posesivních adjektiv (*otcův* a *matčin* – jediných v celém slovníku), komparativních a superlativních tvarů (adj. a adv.) nebo jmenných tvarů adj. / n-ových participií (*nemocen*)

Ne zcela pochopitelný je klíč výběru homonym (homografů) a polysémním lexémů, a to především tehdy, když se neliší jejich segmentace (ani nevyznačená slovnědruhovatost). Podle mého zjištění je zachyceno 782 homonymních nebo polysémních jednotek (1596 vč. opakování), které se v 157 případech liší prvním písmenem (apelativa × propria: např. *kanada* × *Kanada*), jen u 51 se neshoduje segmentace (*pal-i-č-k-a* × *pal-ič-k-a*). Celkem 30 lexémů (lexií) má trojí výskyt (*bosák, klika, box* aj.), 2 dokonce čtverý (*pas* a *kuba/Kuba*). Propria se v celém hesláři vyskytují jen ilustračně (podobně jako např. v SOKOLOVÁ – MOŠKO – ŠIMON – BENKO 1999), v počtu 1105. Na 58 979 nesložených slov připadá 4657 kompozit, to je poměr 92,7 ku 7,3 %. Oddělování afixů, které ztratily svou (sémantickou a funkční) samostatnost, bylo formálně naznačeno závorkováním (pře-/s/věd/č/-i-t), celkem se to týká 2132 případů. Podrobnější kvantitativní zhodnocení slovníku uvádím v kap. 7.

Číselný index uváděný u každé lexikální jednotky má tuto strukturu (přípustné jsou až ternární kombinace, např. 3,5,6 huč-k-a):

- 0 slova autosémantická domácího původu
- 1 slova synsémantická
- 2 slova složená z více kořenových morfémů
- 3 slova autosémantická cizího původu (přejatá)
- 4 vlastní jména
- 5 slova expresivní povahy
- 6 homonymie kořenových morfémů nebo slov
- 7 eponyma (apelativizovaná propria)
- 8 archaická slova
- 9 slova nejasného původu, resp. s nejasnou morfematickou strukturou

Doprovodné seznamy morfémů přinášejí jak informace odvoditelné ze slovníkové části (přehled nejfrekventovanějších afixů, rejstřík prefixálních n-tic apod.), tak i údaje nerekonstruovatelné. Týká se to především registru stylisticky neutrálních kořenových morfémů, uspořádaných do 1931 množin alomorfů a variant (číslováno je jen 1737, zbývající část tvoří sémanticky blízké množiny rozlišené písmeny *a* až *h*), u nichž je arabskými číslicemi rozlišena homonymie, resp. homografie (např. 982 pes, ps<sub>-1</sub> × 997 pís-, pís-, ps<sub>-2</sub>). O podrobnosti výchozího členění v počítačové bázi, z něhož se do hlavního slovníku nepromítlo kromě indexu téměř nic, svědčí např. rozlišování sufixálních morfémů a některých radixů soustavou příznaků:

PLT pluralia tantum, ADJ adjektiva, VB verba, ADV adverbia, SNS synsémantika, TSG transgresiv, propr. = kořenový morfém propria (např. *slov-* v slovenský), epon. = radix eponyma (např. *kub-* v kuba, doutník z Kuby), adv. = adverbium (*líp* od lépe × *líp-a*) nebo akr. zkratkové výrazy (např. *mol* pro grammolekulu).

Pokud by se bývalo podařilo zachránit původní počítačový formát dat, tak jako třeba v případě *Korpusu věcného stylu* neboli *Českého akademického korpusu* (podrobně HLADKÁ – KRÁLÍK 2006), byla by nejen pro mě možnost navázat na přerušenu kontinuitu morfematického výzkum češtiny mnohem snazší. Takto jsem byl nucen odvodit elektronickou podobu díla z tištěné předlohy a další zhodnocování jeho obsahu vést z praktických důvodů trochu jiným způsobem (viz kap. 7 a 8). Digitální verze RMSČ je nyní (bez doplňující anotace popsané v 7.1) přístupná také na internetu<sup>62</sup>.

1	jak-ý-koliv	7	koc-our-k-ov	3,6	kruci-fix
1	kter-ý-koliv	4	Žižkov	1,6	kruci-fix
0	prů-li-v	0	lov	3	in-fix
0	vliv	2	myš-i-lov	3	/su/fix
3,6	mas-iv	2	ryb-o-lov	3	farynx
3	da-t-iv	0,6	do-slov	3,6	box
3	neg-at-iv	0,6	pro-slov	3,6	box
3	in-dik-at-iv	0	vý-lov	3,6	box
3	vok-at-iv	0	dom-ov	3	para-dox
3	/ab/lat-iv	4	Janov	3	flox
3	/super/lat-iv	8	žernov	3,6	lux
3	reg/ul/-at-iv	0	rov	3,6	lux
3	norm-at-iv	0	kr/ov/	4	Benelux
3	nomin-at-iv	8	zá-kr/ov/	3	onyx
3	/kom/par-at-iv	0	pří-kr/ov/	1,6	by
3	imper-at-iv	0	po-kr/ov/	1,6	by
3,6	re-cit-at-iv	0	ostrov	1	aby
3,6	opt-at-iv	0	hřb-it-ov	3,2	elektro-po-třeb-y
3,6	stat-iv	4	Lvov	3,2	foto-po-třeb-y
3	prezerv-at-iv	0	vy-haz-ov	3	ragby
3	fix-at-iv	0	červ	1	jak-by
3	akuz-at-iv	3	nerv	3,2	kata-komb-y
3	a-kred-it-iv	1	teprv	1	rád-o-by
3	prim-it-iv	0	nej-prv	1	jak-o-by
3	gen-it-iv	0	po-nej-prv	0	drob-y
3	/in/fin-it-iv	0	/ob/u-v	0	útr/ob/-y
3	aper-it-iv	0	proti-mluv	0	z-/pů/sob-y
3	poz/it/-iv	0	se-su-v	7	derby
3	dia-poz/it/-iv	4	Vesuv	0	zá-snub-y
3	akt-iv	3	inter-view	0	na-rub-y
3	/ob/jekt-iv	0	kyv	0	/ot/rub-y
3,2	tele-/ob/jekt-iv	0	vý-kyv	1	kdyby
3	/ko/lek-t-iv	0	roz-kyv	0	bez-po-chyb-y
3	korekt-iv	0	od-ry-v	9	do-hromad-y
3	/de/tekt-iv	0	po-ry-v	7	kanad-y
3	kon-junk-t-iv	3	klimax	1,5	jin-ad-y
3	motiv	7	minimax	1,8	on-ad-y
1,2	jak-tě-ži-v	3	borax	0	vnad-y
1,3	jak-ži-v	3,2	pneumo-torax	4	Poděbrad-y

Obr. 8: Ukázka slovníku (RMSČ/SLAVÍČKOVÁ 1975: 412)

<sup>62</sup> <<http://www.morfemy.cz/>>

165	dás/eň/	176	desk- dest-	194	dlah- dlaž- dlažd- dlažd- dláž- dlážd-
166	dáv- <sub>2</sub>	177	dešt- dešt- dešt	195	dlaň- dlaň
167	-db-	178	devat- devát- devět- devít-	196	dlouž- <sub>2</sub> dluh -dluž- <sub>2</sub>
168	dcer- dceř-	179	děc- -dět- -dět- dít-	197	dm- dmuch- dmych- dou- du-
169	dehet -deht-	180	-děč- -děk- <sub>1</sub> -dík- <sub>1</sub>	198	dn- <sub>3</sub>
170	-dech- -dch- -dš- -dych- -dých- -dyš- -douš-	181	děd-	199	dob- <sub>1</sub>
170a	duch <sub>1</sub> duš- <sub>1</sub>				
170b	dycht-				
171	del- dél-				

Obr. 9: Ukázka seznamu kořenových morfémů (RMSČ/SLAVÍČKOVÁ 1975: 518)

#### 5.2.1.2 Šiška: *Bázový morfematický slovník* (2005)

Přesně 30 let po slovníku Slavičkové na knižním trhu objevilo druhé vydání (<sup>1</sup>1998) edukativně zaměřené příručky slovníkového typu, kterou připravil vysokoškolský pedagog Z. Šiška (\* 1930). 228 stran publikace obsahuje předmluvu, poznámku k druhému vydání, studii *K problematice morfematické segmentace slova v češtině* (11 str.), přehled „aplikovaných symbolů, indexů a zkratk“, „výklad fonologické transkripce kořenových morfů a segmentací“, kvantitativní přehled „zastoupení kořenových morfémů podle počtu jejich morfů a segmentací“, *Bázový morfematický slovník češtiny* (90 str.), *Abecední seznam segmentovaných slov a slovních tvarů* (111 str.) a bibliografii. V jazykovědné bohemistice si Šiškovu dílo své místo hledá jen velmi obtížně. Nedávná oborová syntéza (PLESKALOVÁ ET AL. 2007), v níž jsou podle anotace „představeny a charakterizovány všechny rukopisné a pozdější tištěné práce, které měly a mají význam (...)“, o ní (ani o autorovi) není žádná zmínka.

Do slovníku byla pojata vybraná slovní zásoba *Akademických pravidel pravopisu* a *Slovníku spisovného jazyka českého* v rozsahu instruktivní demonstrace představovaného typu morfematické analýzy. Z větší části podle popsání zásad byly lexikální jednotky segmentovány na morfy<sup>63</sup> a následně seříděny podle tzv. báze, kterou tvoří (usouzeno implicate) všechny fonologické podoby alomorfů a variant vyčleněných kořenů. Zahrnuti jsou i gramaticky podmíněné (pádové) variace nemající slovotvornou obdobu (např. /ostrovj/ z Lsg *barvě* nebo /pstruz/ z Npl *pstruzi*), bez přesnějšího odkazu je v bázi vyznačována homomorfie. Podle mého zjištění slovník sestává z 8115 lemmat a 148 slovních tvarů (ty zůstávají nesegmentovány, slouží pouze jako motivanty některých bazových exponentů), které jsou uspořádány do celkového počtu 800 bází (číslování končí hodnotou 745, zbylých 55 je formálním náznakem přiřčeno k jiným bázím: např. 728.1 {zvuk / zvuč} k 728 {zvon / zvon' / zva:n'}. Počet delimitovaných morfémů (morfů) je 34 326, průměrně tedy na jeden lexém (lemma) připadá asi 4,2 morfému (morfu). Jedné bázi průměrně odpovídá 10,1 lexikálních jednotek. Z hlediska počtu morf(ém)ů v lexému jsou ve slovníku (v souladu s uplatněnou koncepcí) 2 lexikální jednotky složené z 1 morf(ém)u (*alibi*, *snáz*), 467 ze 2, 1726 ze 3, 2957 ze 4, 1711 z 5, 1002 ze 6, 222 ze 7, 26 z 8, 5 z 9 a 1 z 10 morf(ém)ů.

<sup>63</sup> Morfém je chápán jen jako abstrakce mající i v systému realizaci výlučně v podobě (alo)morfu (srov. 4.3.4).

V pozadí těchto charakteristik<sup>64</sup> je však nutné vidět poněkud nezvyklé chápání nulových morf(ém)ů (viz 4.3.5.1). Zatímco u slovesných a adjektivních<sup>65</sup> tvarů Šiškovo pojetí vesměs odpovídá koncepci MČ2, u deverbativních substantiv (verbum je ze slovotvorného hlediska automaticky chápáno za prius, substantivum za posterior) počítá nezvykle s nulovou kmenotvornou příponou – např. ve slově vod-o<měr-Ø-Ø –, takže v Nsg (lemmatu) operuje se dvěma nulami. Tento přístup je však konsekvantní jen zdánlivě. Z etymologického úhlu pohledu, který je při morfematické analýze legitimní (viz 6.1), by tak nulové téma muselo být např. i ve slově (mě-Ø-s)-íc (REJZEK 2001). Počítat v povrchové struktuře s *nesoustavnými relikty předhistorické tematické flexe* (Komárek) jinde než u obvyklých typů neuter (jehně, téma) a životných maskulin (kníže) je přinejmenším nepraktické. Celkový pohled na vnitřní strukturu českých slov, tak jak jej předkládá Šiškův slovník, je však zkreslující i jinak. Reprezentativnost omezeného rozsahu slovní zásoby vůči vzájemnému poměru zahrnutých slovních druhů (např. jen 10 zájmen), zastoupení proprií<sup>66</sup> (53) a apelativ (8062) nebo třeba kompozit (228) a nesložených slov (7887) je nepochybně nedostatečná.

Přesto je možné a do jisté míry přínosné prozkoumat alespoň některé vnitřní korelace týkající se tohoto limitovaného souboru. Nabízí se např. srovnání délky (reprezentativních tvarů) lexikálních jednotek vyjádřené v morf(ém)ech a slabikách. Slovník obsahuje 300 lemmat tvořených jednou slabikou<sup>67</sup>, 2425 sestávajících ze 2 slabik, 3539 ze 3, 1570 ze 4, 250 z 5, 28 ze 6, 6 ze 7 a 1 z 8. Průměr na 1 lexém je tedy asi 2,9 slabiky, resp. 4,2 mor(ém)u. Z jiného pohledu, na 1 mor(ém) připadá přibližně 0,69 slabiky, na 1 slabiku 1,46 mor(ém)u. Poměr základních jednotek zvukové (fonologické) a morfologické (morfematické) roviny jazyka je očividně asymetrický, neplatí tak poznatek ze srovnání jiných jazykových plánů, kdy např. hranice morfematická je i zároveň hranicí slovotvornou. Nic na tom nemění známá skutečnost, že stanovení přesných hranic slabiky (tj. rozčlenění fonemického kontinua mezi slabičnými vrcholy ve prospěch kody předcházející a prétury následující sylaby) je mnohdy značně nesnadné, ne-li nemožné.

*Bázový morfematický slovník češtiny* má oficiálně pouze tištěnou podobu, elektronická verze díla byla pořízena v raných fázích promýšlení materiálového zázemí přítomné práce a je nyní přístupná na internetu<sup>68</sup>.

<sup>64</sup> Srovnatelná šetření provedená nad materiálem slovníku Slavičkové (RMSČ) uvádím v kap. 7.

<sup>65</sup> V protikladu k pojetí Komárkovu (KOMÁREK 2006g), je tak finální (koncovkový) morf(ém) -ý/-í (např. v Nsg) u měkkých i tvrdých adjektivních typů považován za téma (kmenotvornou příponu) a tzv. koncovka je nulová (Komárek podává přesvědčivé argumenty pro opačné řešení).

<sup>66</sup> To je však obecnější problém morfematických a morfematicko-derivačních slovníků.

<sup>67</sup> Uvažovat ve flektivním jazyce o n-slabičných lexémech postrádá jakýkoli smysl (např. *snaha* /2slab./, *snah* /1slab./, *snahami* /3slab./ atd.). Kvantifikace morf(ém)ů naopak význam má, protože jejich počet se až na drobné výjimky ((po-)nes-Ø-Ø) při flexi nemění (např. *snah-a* /2morf./, *snah-Ø* /2morf./, *snah-ami* /2morf./ atd.).

<sup>68</sup> <<http://www.morfemy.cz/>>

681	[ vař / va:ř / var / va:r / vř <sup>H</sup> / vr ]	692
	vař-i<t, na-vař-i<t, po-vař-i<t, u-vař-i<t, za-vař-i<t, vař-i<č-Ø, u-vař-Ø<en-ý<Ø, za-vař-Ø<en-in-a, pře-vař-ov<a<t, s-vař-ov<a<t, s-vař-e<t, s-vař-e<č-Ø, s-vař-e<n-i<Ø, za-vař-k-a, var-n-ý<Ø, var-n-a, vár-k-a, vár-k-ov-ý<Ø, vř-i<t, vř-e<l-ý<Ø, vř-e<l-ost-Ø, vř-e<l-e«, vy-vř-i<t, vy-vř-e<l-in-a, vr-ou<c-i<Ø	
681.1	[ vjer <sup>H</sup> ]	696
	vy-věr-a<t, vy-věr-á<n-i<Ø	
682	[ va:z <sup>H</sup> / vi:z / vaz / vjez ]	687
	váz-a<t, na-váz-a<t, o-váz-a<t, pod-váz-a<t, pře-váz-a<t, s-váz-a<t, u-váz-a<t, za-váz-a<t, od-váz-a<n-ý<Ø, pře-váz-a<n-ý<Ø, váz-á<n-i<Ø, váz-nou<t, u-váz-nou<t, u-víz-nou<t, s-vaz-ov<a<t, po-roz-vaz-ov<a<t, po-s-vaz-ov<a<t, vaz-a<dl-o, vaz-a<č-Ø, vaz-b-a, vaz-iv-o, pod-vaz-(e)k-Ø, s-vaz-(e)k-Ø, s-vaz-e<č-(e)k-Ø, ú-vaz-(e)k-Ø, věz-e<t, věz-n-i<t, věz-(e)ň-Ø, věz-n-ic-e	
683	[ včel ]	
	včel-a, včel-k-a, včel-i<č-k-a, včel-stv-o, včel-i<Ø, včel-ař-Ø, včel-ař-sk-ý<Ø, včel-o<mor-k-a	
684	[ vej / vaj ]	
	vej-c-e, vej-c-ov<it-ý<Ø, vej-č-it-ý<Ø, vej-č-it-ě«, vaj-i<č-k-o, vaj-i<č-k-ov-ý<Ø, vaj-e<č-n-ý<Ø, vaj-e<č-n-ik-Ø, vaj-e<č-n-ik-ov-ý<Ø	
685	[ vesel <sup>H</sup> ]	686
	vesel-ý<Ø, vesel-e«, z-vesel-a«, vesel-ost-Ø, vesel-i<Ø, vesel-i<t se, ob-vesel-i<t, roz-vesel-i<t, roz-vesel-Ø<en-ý<Ø, vesel-ic-e, vesel-k-a, ob-vesel-ov<a<t	

Obr. 10: Ukázka slovníku (Šiška 2005: 102)

## 5.2.2 Slovenština

### 5.2.2.1 Morfematický slovník slovenčiny (1999)

Kolektivní *Morfematický slovník slovenčiny* vznikl v letech 1994 až 1998. Autorkou teoretických zásad morfematické analýzy<sup>69</sup> (jsou součástí slovníku) je M. Sokolová, výchozí lexikální materiál byl shromážděn na základě *Pravidiel slovenského pravopisu a Krátkého slovníka slovenského jazyka*. Prvotní segmentace, která byla dále podrobena diskuzi, vycházela z ověřování ve výkladových i etymologických slovnících. Do hesláře byly systematicky doplňovány vidové protějšky (sloves), propria jen v rozsahu *Pravidiel*. Celkem obsahuje slovník 66 469 lexémů (při expanzi *sa* u sloves 70 099), kromě toho se ve slovníku vyskytuje 414 víceslovných jednotek. Nemotivovaná slova a deriváty převažují v počtu 56 836, kompozit je 5925. Každá lexikální jednotka je charakterizována dvojmístný číselný indexem, jednotlivé hodnoty znamenají:

Domácí lexémy (0), převzaté (1), kombinace apelativního kořenu a adaptačního afixu (2), pravá kompozita (3), hybridní kombinace převzatého a domácího radixu (4), lexémy s afixoidy (5), propria (6), synchronně nejasná, ale segmentovatelná struktura (7), nesegmentovatelná lexikální jednotka nebo zastřená motivace (8), zkratková a umělá slova (9). Kombinace indexů znamená: převzaté citátové výrazy (1,1), výpůjčky ze slovanských jazyků (0,1), dtto s domácími afixy (0,2), kompozita s homogenními kořeny, ale původem různými afixy (2,3), kompozita s kořeny a afixy podle původu různými (2,4), kombinace převzatého kořene s prefixoidem a domácími sufixy (2,5), nejasná struktura kořenového morfému, částečně segmentovatelná jednotka, ale podle původu různé afixy (2,7), převzaté zkratkové slovo s domácím afixem nebo domácí zkratkové slovo s převzatým afixem (2,9), kompozitum s jedním propriálním morfémem (3,6), derivát s propriálním radixem (6,0).

<sup>69</sup> Většina zásad je probírána v kap. 6.1.

Identifikaci typů jednotlivých morfémů a jejich vzájemných vztahů umožňuje volba těchto (typo)grafických prostředků:

Jisté hranice (spojovník), potenciální (hranaté závorky), hranice mezi morfémem a submorfémem (dvojtečka, např. na-**bi**:j-a:t'); kořenový morfém (**tučně**), synsémantická lexikální jednotka, derivační nebo modifikační morfémy (obyčejným písmem); gramatické morfémy flexe (*kurziva*); rekonstruovaný foném (kulatými závorkami: ob-(v)rat-Ø); neflektivní slovní druhy (levá francouzská uvozovka na konci řetězce); spojovník (nahrazen lomítkem).

Slovník je doplněn statistickým přehledem (autorem je V. Benko), který se týká (1) počtu lexikálních jednotek podle indexů, (2) počtu autosémantik a synsémantik, inventáře (3) kořenových morfémů a jejich nejfrekventovanějších typů, (4) afixů, (5) derivačních morfémů, (6) modifikačních (vidotvorných) morfémů, (7) gramatických morfémů a (8) konektivních a tematických submorfémů. Dílo nikdy nebylo vydáno ani oficiálně zpřístupněno v elektronické podobě.

0 na-dar-om-nic-u«	0 nad-hodn-oc-ov-a:t'	0 nad-mer-n-ost'-ø
0 na-d-a:t' {na-d-áv-a:t'}	{nad-hodn-ot-i:t'}	0 nad-mer-n-y
0 na-d-áv-a:t' {na-d-a:t'}	0 nad-hodn-ot-a	0 nad-mier-a
0 na-d-áv-k-a	0 nad-hodn-ot-i:t'	0 nad-mier-u«
0 nad-beh-n-ú:t' (si)	{nad-hodn-oc-ov-a:t'}	0 nad-mor-sk-y
{nad-bieh-a:t'}	0 nad-hviezd-n-y	0 na-dm-ú:t' sa {na-dým-a:t' sa}
0 nad-bieh-a:t' (si)	0,8 nádch-a	0 nad-ná-rod-n-ost'-ø
{nad-beh-n-ú:t'}	0 nad-chádz-a:t' {nad-ís-t'}	0 nad-ná-rod-n-y
0 nad-byt-oč-n-e«	0 nad-chádz-k-a	0 nad-náš-a:t' (sa) {nad-nies-t'}
0 nad-byt-oč-n-ost'-ø	0 na-dch-n-ú:t' (sa)	0 na-dnes«
0 nad-byt-oč-n-y	{na-dch-ýn-a:t'}	0 nad-nes-e:n-e«
0 nad-byt-ok-ø	0 nad-chod-ø	0 nad-nes-e:n-ost'-ø
0 nad-čas-ø	0 nad-chod-i:t' {nad-ís-t'}	0 nad-nes-e:n-y
0 nad-čas«	0,7 nádch-ov-y	0 nad-nies-t' (sa) (-nes-ie:ø)
0 nad-čas-ov-o«	0 na-dch-ýn-a:t' (sa)	{nad-náš-a:t'}
0 nad-čas-ov-ost'-ø	{na-dch-n-ú:t'}	2 nad-norm-ál-n-y
0 nad-čas-ov-y	0 na-dia-t' (-de:j-e:ø)	2 nad-norm-at:fv-n-y
0 nad-človek-ø (-Pud-ia)	{na-die-v-a:t'}	0 nado«
0 nad-dlž-i:t' (sa)	0 na-diel'-a:t' {na-del-i:t'}	0 [ná]dob-a
0 nad-dvih-a:t' (sa)	0 ná-diel-k-a	0 [ná]dob-k-a
{nad-dvih-n-ú:t'}	0 na-die-t' (-de:j-e:ø)	0 nad-oblač-n-y
0 nad-dvih-n-ú:t' (sa) {nad-dvih-	{na-die-v-a:t'}	0 nad-obl-ič-k-a
-a:t', nad-dvih-ov-a:t'}	0 na-die-v-a:t' {na-die-t', na-dia-t'}	0 nad-obl-ič-k-ov-y
0 nad-dvih-ov-a:t' (sa)	2 na-dikt-ov-a:t' {dikt-ov-a:t'}	0 na-dobr-o«
{nad-dvih-n-ú:t'}	1 nadir-ø	0 na-do-búd-a:t'
0,1 [ná]dej-ø	0 nad-ís-t' (-íd-e:ø, -iš-ic:l-ø)	{na-do-bud-n-ú:t'}
0,2 [ná]dej-a:t' sa	{nad-chádz-a:t', nad-chod-i:t'}	0 na-do-búd-a:tel'-ø
0,2 [ná]dej-n-e«	0 na-dív-a:t' sa	0 na-do-búd-a:tel'-k-a
0,2 [ná]dej-n-ost'-ø	0 na-div-o«	0 na-do-búd-a:tel'-sk-y
0,2 [ná]dej-n-y	0 na-div-o«	0 na-do-bud-n-ú:t'
2 na-[de]krét-ov-a:t'	0 na-div-ok-o«	{na-do-búd-a:t'}
0 na-del-i:t' {na-diel'-a:t',	0 na-div-ok-o«	0 nad-oč-n-ic-ov-y
na-del'-ov-a:t'}	0 nad-jazd-ø	0 na-doj-i:t'
0 na-del'-ov-a:t' {na-del-i:t'}	0 nad-klad-a:t' {nad-lož-i:t'}	0 na-dol«

Obr. 11: Ukázka slovníku (Sokolová – Moško – Šimon – Benko 1999: 224)

#### 5.2.2.2 Slovník kořenových morfém slovenčiny (2005)

Slovník představuje další zpracování obsahu *Morfematického slovníku slovenčiny*, autorsky se na něm podílel (v letech 2003 a 2004) pozměněný kolektiv vedený opět M. Sokolovou. Metodologická východiska a zásady zpracování byly s předstihem publikovány časopisecky (SOKOLOVÁ – IVANOVÁ – DZIAKOVÁ 2003). Beze změn bylo převzato členění jednotek, způsob notace i systém číselných indexů. Novinkou je (i vícenásobné) odsazení jednotek pomocí symbolu „>“ naznačující jednotlivé derivační kroky. Práce má morfematicko-derivační charakter, který se projevuje v uspořádání hesel slovníku podle kořenů. Většinou se toto členění shoduje se slovotvornými hnízdy, u polysémnních radixů však může vzniknout rozsáhlejší soubor. Výchozí počet hesel byl snížen (ze 70 099 na cca 66 500) rozhodnutím neopakovat polysémnní lexie a optimalizovat zachycení reflexivity (sloučením *sa*, *si*). Celkem



je rozlišeno 7964 radixů, 825 z nich je homonymních, 655 homonymních se dvěma významy. Slovníková část je členěna na *apelativní část* (obsahující všechna autosémantika včetně těch bez motivace, a synsémantika motivující i motivovaná), *propriální část* (soubor s propriálními motivanty) a samostatný celkový přehled *synsémantik*. Ani toto dílo nebylo oficiálně zpřístupněno v jiné než tištěné podobě.

0 > čo hned«	0 > ob-hni-ť {ob-hní-v-a:ť}	HOBEL(HOBL)
HNEĎ <sup>2</sup> / HNED	0 > ob-hní-v-a:ť {ob-hní-ť}	1 hobel'-ø
HNET / HNIES	0 > pod-hni-ť {pod-hní-v-a:ť}	2 > hobl-ik-ø
HNEV	0 > pod-hní-v-a:ť {pod-hní-ť}	2 >> hobl-ič-ek-ø
0 hnev-a:ť {na-hnev-a:ť}	0 > pre-hni-ť {pre-hní-v-a:ť}	2 >> hobl-ik-ov-ý
0 na-hnev-a:ť {hnev-a:ť}	0 > pre-hní-v-a:ť {pre-hní-ť}	2 > hobl'-ov-a:ť {o-hobl'-ov-a:ť}
0 << hnev-ø	0 >> pre-hni-t-ý	2 > o-hobl'-ov-a:ť {hobl'-ov-a:ť}
0 > hnev-ník-ø	0 >>> pre-hni-t-ost'-ø	2 >> hobl-in-a
0 >> hnev-nič-k-a	0 > za-hni-v-a:ť	2 >>> hobl-in-k-a
0 > hnev-liv-ý	HŇÍ / HŇI	2 >> hobl'-ov-ač:k-a
0 >> hnev-liv-ost'-ø	HNID	2 >> hobl-ic-a
0 >> hnev-liv-o«	0 hnid-a	2 >> hobl'-ov-ac-i
0 > na-hnev-a:n-ý	HNIES (HNET)	2 >> z-hobl'-ov-a:ť
0 >> na-hnev-a:n-ost'-ø	0 hnies-ť	HOBL / HOBEL
0 >> na-hnev-a:n-e«	0 > hnet-ac-i	HOBOJ
0 >> na-hnev-a:n-o«	0 >(>) hnet-ač-ø	1 hobj-ø
0 > hnev-a:ť sa	HNIEZD (HNIEZĎ)	1 > hobj-ist-a
{na-hnev-a:ť, po-hnev-a:ť}	0 hniezd-o	2 >> hobj-ist-k-a
0 > na-hnev-a:ť sa {hnev-a:ť}	0 > hniezd-i:ť (sa)	2 > hobj-ov-ý
0 > po-hnev-a:ť sa {hnev-a:ť}	0 >> hniezd-iev-a:ť	HOBR
0 > do-hnev-a:ť	0 >> po-hniezd-i:ť sa	1,9 hobr-a
0 > roz-hnev-a:ť (sa)	0 >> u-hniezd-i:ť sa	2,9 > hobr-ov-ý
0 >> roz-hnev-a:n-ý	0 >> za-hniezd-i:ť sa	HOBY / HOBBY
0 >>> roz-hnev-a:n-e«	{za-hniezd'-ov-a:ť}	HOD <sup>1</sup>
0 >>> roz-hnev-a:n-o«	0 >> za-hniezd'-ov-a:ť sa	0 hod-y
0 > po-hnev-a:ť si	{za-hniezd-i:ť}	0 > hod-ov-ý
HŇÍ (HŇÍ, HNO)	0 > hniezd-oč:k-o	0 << hod-ov-a:ť
0 hni-ť (hni:j-e:ø)	0 > hniezd-n-y	0 >> hod-ov-ník-ø
{po-hni-ť / z-hni-ť}	0 > hniezd-ov-ý	0 >>> hod-ov-nič-k-a
0 po-hni-ť / z-hni-ť {hni-ť}	HNIEZĎ / HNIEZD	0 >>> hod-ov-nic-k-y
0 > hno-j-ø	HNIS	0 >> hod-ov-n-ý
0 >> hno-j-i:ť {po-hno-j-i:ť}	0 hnis-a:ť {z-hnis-a:ť}	HOD <sup>2</sup> / HOD <sup>1</sup>
0 >> po-hno-j-i:ť {hno-j-i:ť}	0 z-hnis-a:ť {hnis-a:ť}	HOD <sup>3</sup> / HOD <sup>2</sup>
0 >> hno-j-isk-o	0 << hnis-ø	HOĎ <sup>1</sup> (HÁDZ, HÁDZ, HÁDŽ,
0 >> hno-j-n-ý	0 >> hnis-ov-ý	HOD <sup>2</sup> )
0 >>> hno-j-n-ic-a	0 > hnis-av-ý	0 hod-i:ť (sa) {hád-z-a:ť}
0 >>>> hno-j-n-ič-n-ý	0 >> hnis-av-ost'-ø	0 hádz-a:ť (sa) (-hádž-e:ø)
0 >>>> hno-j-n-ic-ov-ý	HNO / HŇI	{hod-i:ť}
0 >> hno-j-ov-ý	HNUS	0 > hod-ø
0 >>(>) hno-j-ov-ic-a	0 hnus-ø	0 > hádz-a:n-á
0 >>>> hno-j-ov-ic-ov-ý	0 << hnus-i:ť {za-hnus-i:ť}	

Obr. 12: Ukázka slovníku (SOKOLOVÁ – OLOŠTIK – IVANOVÁ 2005: 159)

## 5.2.3 Polština

### 5.2.3.1 Słownik gniazd słowotwórczych współczesnego języka ogólnopolskiego (2001–2004)

Slovotvorný výzkum polštiny zaměřený na budoucí slovníkové zpracování probíhal s různou intenzitou od roku 1986. Nakonec čtyřdílný hnízdový slovník byl vydán tiskem počátkem 21. století s obsahem jednotlivých částí takto: /1/ Gniazda odprzymiotnikowe (základem adjektiva), /2/ Gniazda odreczownikowe (substantiva), /3/ Gniazda odczasownikowe (slovesa), /4/ Gniazda motywowane przez liczebniki, przysłówki, zaimki, przymyki, modulanty, onomatopeje, wykrzykniki (ostatní slovní druhy). Část slovníku byla sestavena poloautomaticky s využitím počítačové techniky, ruční zpracování a úpravy si vyžádala zejména složitá sémantická stránka. Zahrnutá slovní zásoba je opatřena důmyslnou soustavou kvalifikátorů (např. údaje o produktivitě, oborovém zařazení nebo o homonymii). V případě potřeby jsou doplněny definice významu. U každé jednotky je symbolickým zápisem a/nebo odkazem na motivující slovo zachycen postup, kterým byla utvořena.

OKNO 2., 3. – b. der.	
[OKNO] 3. 'otwór, dziura w czym (czasem zamykane lub zabijane)'	
[okien-ko] 3. *, 6. * <i>bot.</i> , 7. * <i>lotn.</i>	S,S
[OKNO] 6. <i>górn.</i> , 7. <i>low.</i> , 8. <i>techn.</i> – b. der.	
OKO 1. 'narząd wzroku'	
[jedn-o-ocz-(e)] Supl. <i>med.</i>	(N+)S,S
mon/ok-] 1. 'szkło optyczne na jedno oko'	(O+)S,S
[monokl-owy]	(O+)S,S,Ad
[oc-(o)]	S,S
ocz-ęta	S,S
ocz-ko 1.	S,S
[ocz-ąt-ko]	S,S,S
ocz-k-(ować) 1. <i>pot. żart.</i>	S,S,V
ocz-lik* <i>zool.</i>	S,S
[ocz-nik] 2. <i>techn.</i>	S,S
ocz-o-dół	(S+)S,S
[nad   oczodoł-owy] <i>anat.</i>	(S+)S,S(+P),Ad
[oczodoł-owy]	(S+)S,S,Ad
[ocz-o-pląs]* <i>med.</i>	(S+)S,S
[ocz-yisko]	S,S
ok-a-mgnienie	(S+)S,S
ok-ular 1. <i>fiz.</i>	S,S
[bin/okular] Supl. <i>fiz.</i>	S,S(+O),S
[binokular-ny] Supl. <i>fiz.</i>	S,S(+O),S,Ad
ok-ulary 1.	S,S
[okular-ki] Supl.	S,S,S
okular-nik* <i>zool.</i>	S,S,S
[okular-nik] Supl. 3. <i>pot. żart.</i>	S,S,S
[okularz-yiska]	S,S,S
[okular-owy] 1.	S,S,Ad
ok-ulary 2.	S,S

Obr. 13: Ukázka slovníku (JADACKA ET AL. 2001–2004, Tom 2, s. 955)

## 5.2.4 Ruština

### 5.2.4.1 Russian Derivational Dictionary (1970)

Jeden z prvních slovníků vznikajících za pomoci počítačové techniky byl připravován od roku 1962 ve Spojených státech díky podpoře amerického letectva. Ruský lexikální materiál o velikosti 110 tisíc slov byl (kvůli obavám ze selhání techniky dvojmo) zpracován na děrných štítcích převáděných v dalších fázích na magnetické pásky. Po pečlivém zvážení všech výhod i rizik byla zvolena kombinovaná metoda segmentace, při níž bylo s využitím speciálně sestaveného algoritmu nejprve provedeno automatické rozčlenění slov na morfémy (údajně trvalo za asistence několika lidských operátorů zhruba dva a půl měsíce). Následovala ruční korektura. Vyčleňovány byly i segmenty bez přímého vztahu k významu, celkem bylo potřeba rektifikovat asi 27 % morfematických hranic (potíže činily převážně homografy). Segmentovaná slova byla dále setříděna podle kořenů s přihlédnutím k problematice kompozit (odlišení derivace a kompozice). Při sestavování jednotlivých slovních hnízd se některé sporné případy ověřovaly pomocí slovníků. V konečné podobě má slovník 10 953 hnízd a v závislosti na alomorfii 1–16 podhnízd.

ВОР 1	ОБ ВОР ОВ ЫВА ТЬ ОБ ВОР ОВ ЫВА ТЬ СЯ ПС ВОР ОВ А НН ЫИ ПС ВОР ОВ А ТЬ ПС ВОР ОВ ЫВА ТЬ ПРС ВОР ОВ А ТЬ СЯ ПРО ВОР ОВ ЫВА ТЬ СЯ РАЗ ВОР ОВ А НН ЫИ РАЗ ВОР ОВ А ТЬ РАЗ ВОР ОВ ЫВА ТЬ РАЗ ВОР ОВ ЫВА ТЬ СЯ С ВОР ОВ А НН ЫИ С ВОР ОВ А ТЬ У ВОР ОВ А НН ЫИ У ВОР ОВ А ТЬ	ВОРОЖ	НА ВОРОЖ Е НН ЫИ НА ВОРОЖ И ТЬ ОБ ВОРОЖ Е НН ЫИ ОБ ВОРОЖ И ТЕЛ Ь Н ОСТ Ь ОБ ВОРОЖ И ТЕЛ Ь Н ЫИ ОБ ВОРОЖ И ТЬ ОБ ВОРОЖ И ТЬ СЯ ПО ВОРОЖ И ТЬ ПРИ ВОРОЖ Е НН ЫИ ПРИ ВОРОЖ И ТЬ ПРО ВОРОЖ И ТЬ ВРАГ ВРАЖ Е СК ИИ ВРАЖ ИИ ВРАЖА А ВРАЖА ЕБ Н ОСТ Ь ВРАЖА ЕБ Н ЫИ ВРАЖА ЕБ Н ЫИ ВРАЖА ОВ А ТЬ ЗА ВРАЖА ОВ А ТЬ
ВОР 2	С ВОР А С ВОР И ТЬ	КЛАСС ОВ О	
ВОРВАН	ВОРВАН Ь		

Obr. 14: Ukázka slovníku (WORTH – KOZAK – JOHNSON 1970: 101)

#### 5.2.4.2 Slovar' morfem russkogo jazyka (1986)

Slovní zásoba slovníku, jejíž celkový rozsah je nakonec 52 tisíc slov, prošla několika fázemi vyvažování (zejm. z hlediska proporcionálního rozvrstvení slovní zásoby). Vyloučena byla např. mezinárodní slovní zásoba, „barbarismy“ a překvapivě také zájmena, pokud nefundují jiné slovní druhy (substantiva, adjektiva a verba). Celkový počet morfémů, které byly delimitovány z uvedeného množství hesel, je asi 5000. Z toho připadá více než 4400 na kořeny (1800 z nich je „alomorfních“), 70 na prefixy a přibližně 500 na sufixy, mezi nimiž se vyskytlo mnoho „anomálních“. Kniha je členěna na úvodní teoretické expozé, slovníkovou část a přílohy. Jádro slovníku má čtyři části: kořenovou, prefixovou, sufixovou a index. V přílohách jsou uvedeny rejstříky radixů, homonymních kořenů, prefixů, sufixů, nejčastějších sufixů a prefixů, nejproduktivnějších modelů vnitřní struktury slov a nejběžnějších slovotvorně produktivních morfematických kombinací. Slovník umožňuje mnohostranný vhled do morfematické struktury ruštiny. Díky odkazovému indexu (generálnímu hesláři) a „principu relační databáze“ (tj. tabulky se záznamy propojenými tzv. primárními a cizími klíči) je možné vyhledat např. všechny odvozeniny od kořenu hledaného slova. Z hlediska cíle zachytit složité vztahy typu M:N v tištěné formě je to zřejmě nejefektivnější řešení.

<b>РЕХ</b>	раз-√-а-ть <sup>1</sup> , -(ть-ся) <sup>2</sup>
√-нў-ть-ся	раз-√-ёни-е
<b>1 РЕЧ (1 рек)</b>	раз-√-й-ть <sup>1</sup> , -(ть-ся) <sup>2</sup>
√-еньк-а	раз-√-й-м-ый
√-к-а	раз-√-й-тел-ь-н-ый
√-н-ой	<b>РЕШЁТ (решёт, решёч)</b>
√-н-ик-е	√-ё
√-онк-а	√-й-ть
√-ушк-а	√-ин-а
за-√-бе	√-н-ик-е
за-√-н-ый	√-ц-ё
по-√-бе	√-чат-ый*
при-√-бе	из-√-й-ть <sup>1</sup> , -(ть-ся) <sup>2</sup>
при-√-н-ый	об-√-н-ть <sup>1</sup> , -(ть-ся) <sup>2</sup>
<b>2 РЕЧ (2 рек, риц, рók, рóч, 2 рóш)</b>	об-√-ин-а
√-бё	под-√-н-ть
√-ев-ой	под-√-ин-а
√-ев-йк-е	под-√-н-ик-е
√-ёни-е	про-√-н-ть
√-ённ-ый	<b>РЕШЁТ (решёт, решёч)</b>
√-йв-ый	√-к-а
√-йв-ост-бё	√-н-ый
√-йст-ый	√-оч-н-ый
√-йст-ост-бё	√-ц-е
за-√-бё-ся	√-чат-ый*
из-√-бё	об-√-к-а
из-√-ёни-е	<b>РЕШЁЧ (решёт, решёт)</b>
на-√-бё <sup>1</sup> , -(бё-ся) <sup>2</sup>	за-√-енн-ый
на-√-ёни-е	из-√-ива-ть <sup>1</sup> , -(ть-ся) <sup>2</sup>
на-√-ённ-ый	об-√-ива-ть <sup>1</sup> , -(ть-ся) <sup>2</sup>
на-√-и-е	под-√-ива-ть <sup>1</sup> , -(ть-ся) <sup>2</sup>
на-√-н-ый	<b>1 РЖ</b>
об-√-бё <sup>1</sup> , -(бё-ся) <sup>2</sup>	√-а
об-√-ёни-е	√-ав-ый
об-√-ённ-ый	√-ав-е-ть
об-√-ённ-ост-бё	√-ав-н-ть
от-√-бё-ся	√-ав-л-ени-е
от-√-ёни-е	√-ав-ост-бё
от-√-ённ-ый	√-ав-чин-а
пред-√-бё	√-ав-чин-н-ый

наречение — 2 реч  
наречённый — 2 реч  
наречие — 2 реч  
наречный — 2 реч  
наречь — 2 реч  
наречься — 2 реч

Obr. 15: Ukázka slovníku (KUZNĚCOVA – JEFREMOVA 1986: 281)

#### 5.2.4.3 Tichonov: Morfemno-ortografičeskij slovar' russkogo jazyka (1996)

Prakticky zaměřený slovník je výsledkem snahy zkombinovat dva typy lexikonů: ortografický a morfe-matický. Autor vychází z faktu, že zásady ruského pravopisu zahrnují také morfe-matický princip. Podle jeho slov jde o princip hlavní. Dokládá to na příkladu (svého) staršího slovotvorného slovníku o velikosti 145 tisíc slov, v němž je pouze 19 tisíc slov (méně než 8 %) neodvozených (непроизводны). Právě derivate je příčinou značné grafematické variability (zejm. kořenů). Naučit se pravidla (ruského) pravopisu je prakticky nemožné bez dobré znalosti zákonitostí slovotvorby. Tomuto závěru vychází vstříc slovníková část čítající na 95 tisíc ručně analyzovaných slov. To by nebylo možné bez podrobně stanovených zásad. Uživatel slovníku je spolu s podrobným komentářem najde zařazené v rozsáhlé studii nazvané *Morfemika* (česky morfe-matika) zařazené na konec publikace. Mikrostruktura slovníku obsahuje kromě segmentovaných slov údaje týkající se formální morfologie, slovotvorné motivace a také indexy rozlišující polysémnní (homografní) lexémy.

ОТВ	ОТВ	ОТГ
от/влеч/ённ/ость/ от/влеч/ённ/ый; кр. ф. прич. от/влеч/ён/, от/влеч/ен/а; кр. ф. прил. от/влеч/ён/, от/влеч/ённ/а от/влеч/ь(ся) от/вод/ь/ от/вод/и/ть <sup>1</sup> , несов. (от от/вес/ти) от/вод/и/ть <sup>2</sup> , сов. (от вод/и/ть) от/вод/и/ть/ся, несов. (от от/вод/и/ть <sup>1</sup> ) от/вод/к/а от/вод/н/ой и от/вод/н/ый от/вод/ок/ от/воёва/нн/ый от/воёва/ть(ся) от/воёв/ыва/ть(ся)	отвращ/ени/е [й/э] отвращ/ённ/ый от/вык/а/ть от/вык/ну/ть [ср.: при/вык/ну/ть] от/вык/ш/ий от/выч/к/а от/вяз/а/нн/ый от/вяз/а/ть(ся) от/вяз/к/а от/вяз/ыва/ни/е [й/э] от/вяз/ыва/ть(ся) от/гад/а/нн/ый от/гад/а/ть от/гад/к/а от/гад/чик/ от/гад/чик/а от/гад/ыва/ни/е [й/э] от/гад/ыва/ть	от/гранич/ени/е [й/э] от/гранич/енн/ый от/гранич/ива/ть(ся) от/гранич/и/ть от/граф/и/ть от/графл/ённ/ый от/графл/я/ть от/греб/а/ни/е [й/э] от/греб/а/ть(ся) от/греб/ённ/ый от/грёб/к/а от/грёб/ш/ий от/грем/ё/ть от/грес/ти/сь от/грох/а/ть от/грохот/а/ть от/груз/а/ть(ся) от/груз/енн/ый и от/груз/ённ/ый от/груз/и/ть(ся)

Obr. 16: Ukázka slovníku (Tichonov 1996: 347)

#### 5.2.4.4 Chronological Morphemic and Word-Formational Dictionary of Russian

*Chronologický morfemický a slovotvorný slovník ruštiny* se od všech ostatních děl z tohoto přehledu liší hned v několika ohledech. Práce na něm nebyla dosud ukončena a s ohledem na zaměření díla se to ani nepředpokládá. Ač vzniká už od roku 1989, nebyl nikdy vydán tiskem (existuje jen v elektronické podobě). Jeho největší zvláštností je zaznamenávání přibližného data, od kdy se příslušné slovo v ruštině používá (v tom má blízko k velkému *Oxford English Dictionary*). V roce 2000 obsahoval kolem 180 tisíc hesel. Počítačový způsob zpracování umožňuje podrobnou klasifikaci každého lexikálního hesla, což obnáší kromě obligatorního morfemického členění a časového zařazení např. údaje o homonymii, slovním druhu, jazyku původu, počtu významů nebo frekvenci v referenčních zdrojích. Spíše než o slovník jde vlastně o specializovanou lingvistickou databázi. Roztřídění konkrétních záznamů podle sedmi období vývoje jazyka otevírá dosud netušené možnosti zkoumání morfemické struktury ruštiny i v diachronních souvislostech.

Fragment of the main DB of "Chronological Morphemic and Word-Formational Dictionary of Russian"														
Word	Homonymic index of word	Generalised root	Hom. index of gen. root	Part of sp. 1	Part of sp. 2	Pref. 2	Pref. 1	Root 1	Suf. 1	Suf. 2	Suf. 3	Word frequency	Word fixation date	Word age period
ад		ад		сущ				ад				8	1050	5
адажио		адажио		н	сущ			адажио				0	1800	3
адов		ад		п				ад	ов			0	1050	5
адрес		адрес		сущ				адрес				93	1780	3
адресант		адрес		сущ				адрес	ант			0	1866	5
адресантка		адрес		сущ				адрес	ант	к	а	0	1866	3
адресат		адрес		сущ				адрес	ат			4	1891	2
адресный		адрес		п				адрес	н	ый		1	1866	2
адресовать		адрес		гл				адрес	ов	а	ть	11	1780	3

Obr. 17: Ukázka slovníku (Polikarpov 2000)

### 5.2.5 Ukrajinština

#### 5.2.5.1 Poljuha: *Morfemnyj slovar'* (1983)

Slovník obsahuje 36 tisíc hesel spisovné ukrajinštiny, které morfematically člení na prefixy, kořeny, sufixy a zakončení. Abecedně řazený heslář je doplněn údaji o výslovnosti, formální morfologii a možných pravopisných potížích. Za hlavní slovníkovou částí následují samostatné seznamy prefixů, sufixů a kořených morfémů. Z kvantitativního hlediska jde o

90 prefixů, 280 sufixálních morfémů a 1900 vybraných kořenů domácího původu, vždy včetně jejich variant. Kromě těchto registrů je do slovníku pojat přehled 250 nejčastějších slovotvorných komponentů (prefixálních, sufixálních aj. spojení) a 120 slov majících neobvyklou, nepravidelnou strukturu, které jsou rovněž bezprostředně členěny na morfémy. Publikace je explicitně určena lingvistům, učitelům i studentům. V tiráži je uveden náklad 27 tisíc výtisků, což naznačuje, že jde o jakýsi velmi praktický ukrajinský pendant slovníkové části českých pravidel pravopisu, který v přibližně stejném rozsahu (česká PČP mají kolem 34 tisíc hesel) nabízí mnohem hlubší vhled do struktury jazyka.

кравець (~в[ц'/á])	крас/й/ти (краш/ý)
кравець/к/ий (~ць ↔ ськ)	крас/і/ти
кравц/юва/ти	крас/н/ий
кравч/ик	крас/н/о/мóв/н/ий
кравч/йн/а	крас/н/о/мóв/н/ість (~н/ост/і)
кравч/йн/я	крас/от/á
крад/ен/ий	крас/ти (крад/ý, кра/в)
крад/іж (~д/еж/ý)	крас/увá/нн/я
крад/іж/к/а (~ж/ц/і, ~ж/ок)	крас/увá/ти/ся
крадь/к/омá	крас/ýн/я
кра/є/вїд (кра[й/е])	крат/н/ий
кра/є/знá/в/ств/о (кра[й/е])	кращ/а/ти
кра/єч/ок (кра[й/е]ч)	кращ/ий
кра/їн/а (кра[й/і]н)	кра/я/н/ий (кра[й/а]н)
край/к/а (~й/ц/і, ~й/ок)	кра/я/ти (кра[й/а])
край/неб/о	крів/н/ий
край/н/ій	крев/н/як
край/н/ість (~н/ост/і)	кредіт/н/ий
край/н/ощ/і	кредит/óв/ий
край/ов/ій	кредит/óр

Obr. 18: Ukázka slovníku (Poljuha 1983: 144)

#### 5.2.5.2 Karpilovs'ka: Korenevyj gnizdovyj slovnyk ukrajinskoji movi (2002)

Morfematicko-derivační slovník ukrajinštiny zprostředkovává velmi přehlednou formou formální i sémantickou strukturu jednotlivých slov. Lexikon o velikosti 71 304 slov obsáhl značně široké spektrum slovní zásoby včetně nářečního materiálu a cizích slov. Do 653 skupin (hnízd) bylo uspořádáno celkem 2594 kořenů, 1820 z nich je homografních. Na jedno slovotvorné hnízdo tak připadá rozpětí od 2 až do 14 homografních radixů. Slova neobsahujících homografní kořen je jen zhruba 16 %. Z hlediska homonymie a významu připadá na jeden radix v průměru 8 významů, na jeden homonymní radix potom 4 významy. Díky zvolenému řešení grafického rozvržení tisku (použití piktogramů, indexů, značek, odsazení a dalších prostředků) je celkový způsob prezentace složitých fundačně-motivických vztahů (alespoň v menších hnízdech) poměrně instruktivní, i když pochopitelně nemůže konkurovat multidimenzionalitě počítačového přístupu.

<b>КУЛ<sup>2</sup>- 5; куль<sup>2</sup>- 17. Разом – 22</b>				
кул(я) <sup>2</sup> ⇔ (милиция) див. кул <sup>1</sup> -	куль-г(а) ⇔	кульг-áv(ий) ⇔  кульг-áти ⇔	кульгáв-ість кульгáв-іти ⇔  кульгá-ючи кульгá-нн(я) до-кульгáти за-кульгáти на-кульг-увати ⇔ по-кульгáти по-кúльг-увати при-кульг-увати ⇔ про-кульгáти	с-кульгáвіти     накúльгува-нн(я)   прикульг-óв-увати
	кул-яг(а) кул-іти кул-яти ⇔	кул-яв(ий)		
кул <sup>3</sup> - див. КУЛЬ <sup>3</sup> -				
<b>КУЛОН<sup>1</sup>- (від франц. coulant "коштовність" з couler "текти, спадати") – 3</b>				
кулóн-0 <sup>1</sup> (прикраса) ⇔ див. кулон <sup>2</sup> -	кулóн-чик кулóн-н(ий)			
<b>КУЛОН<sup>2</sup>- (від прізвища франц. фізика Ш.- О. Кулона) – 9</b>				
кулóн-0 <sup>2</sup> ⇔ (фізична одиниця) див. кулон <sup>1</sup> -	кулон-о-мётр-іj(а) див. метр <sup>4</sup> - кіло-кулóн мега-# мікро-# мілі-# нан-о-# пік-о-# див. пік <sup>3</sup> - кулóн-метр див. метр <sup>1</sup> -			
куль <sup>1</sup> - див. КУЛ <sup>1</sup> -				
куль <sup>2</sup> - див. КУЛ <sup>2</sup> -				
<b>КУЛЬ<sup>3</sup>- 3; кул<sup>3</sup>- 2. Разом – 5</b>				
куль-0 ⇔	кúл-ик ⇔ куль-ók ⇔	кúлик-ом кульч-ій		

Obr. 19: Ukázka slovníku (Karpilovs'ka 2002: 377)

## 5.2.6 Němčina

### 5.2.6.1 Augst: *Lexikon zur Wortbildung: Morpheminventar* (1975)

Augstův lexikon je typologicky velmi zvláštní druh slovníku. Kombinuje informace z oblasti morfematiky, slovo tvorby, (formální) morfologie, ortografie a v malé míře také některé rysy jednojazyčného výkladového a etymologického slovníku. Makrostruktura je postavena na principu abecedně řazených autosémantických kořenů tvořících jednotlivá slovo tvorná hnízda (Wortfamilien). Jejich počet je přibližně 10 tisíc. Synsémantika jsou zastoupena asi 200 morfémů, konkrétně 52 z nich přísluší flexi, 11 tvoří konektémy (Fugenmorpheme), zbývající počet připadá na afixy. Nahlíženo konceptem volných a vázaných forem (morfémů), zhruba 10 tisíc plnovýznamových lexikálních morfémů je volných (např. *Haus*), kolem 300 patří mezi vázané (*Brom-/beere*/, */Stein/-metz* aj.). U synsémantik existuje asi 200 morfémů, resp. gramatických forem (*auf*, *der*, *wo*), které patří mez volné, k vázaným se řadí 60 morfémů pro flexi (*/Hund/-e*), 11 konektémů (*/Kind/-er/-garten*/) a asi 180 afixů (*/gift/-ig*).

1. KLAPP	1. Schallw.	
	2. V. 'umschlagen, umdrehen'	
	Zus. (-ø)	
	zusammen, um... Abl.	-bar
		-er 1. (S.) f. (-en)
		2. V. Zus.
		-ig
	ge...	-ø Sn. Pl. unz.
		-s Sm. -e; Zus. -ø
	ver-	-ø V. 'veralbern'
		(U) -chen
		(U) -lein
klappe	3. (Sf. -en)	
	Zus. (-en)	
	í... Abl.	-ig
	Vgl. klipp-klapp, vgl. klipp	
	beachte: Abklapper V. 'von Haus zu Haus gehen'	
	úmklapp V. (x) 'umkippen'	
	zusammenklapp V. (x) 'erschöpft sein'	
2. KLAPP	V. 'in Ordnung sein, reibungslos funktionieren'	
	etym. 1 = 2; heute? Die Informanten sind unentschieden.	
KLAR	1. Adj.	
	Zus. (-ø)	
	ún- Abl.	-ø
	er, ver-	U -ø V. Zus.
	úner, uner, er...	-bar
	úner, uner...	-keit
	verún-	-end adjektiv.
		Part. Präs.
	er, áuf...	-er
	áuf...	-isch
	uner, er, úner...	-lich
	uner, úner...	-keit
	únge, áufge...	-et adjektiv.
		Part. Perf.
	áufge...	-heit
	ver, er-	-ung Zus.
	ún-	-heit

Obr. 20: Ukázka slovníku (AUGST 1975, vol. 1, s. 533)

### 5.3 Shrnutí

Uvedený přehled slovníkového zpracování morfematiky a derivatologie pochopitelně není úplný. Zabýval jsem se pouze takovými tituly, jejichž existenci bylo možné dostupnými prostředky empiricky ověřit a které jsou katexochén důležité z hlediska bohemistiky. Vědomě jsem vynechal stručný přehled gramatických morfémů staroslověňštiny (RODOVICH 1971) a etymologicky zaměřené vádemékum distribuce indoevropských kořenů (BIRD 1982), s kterými jsem se sice mohl seznámit a které nepochybně mají bohemikální význam, ale jejichž zaměření nebylo s ostatními díly kompatibilní. Otázky morfematické struktury slov jsou pochopitelně naléhavější a badatelsky atraktivnější v závislosti na typologických charakteristikách jednotlivých jazyků. Kromě slovanských jazyků, u nichž je morfematická segmentace důležitá zejm. kvůli vysoce rozvinuté flexi a bohatství hláskových alternací, panuje velký zájem o morfematickou analýzu jazyků ugrofinských (především finštiny), kde však dominující aglutinační prvek vytváří spíše předpoklady pro formální počítačové zpracování (srov. např. CREUTZ – LAGUS 2002).

Společným rysem představených děl byla snaha využít v rámci omezených možností tištěné publikace univerzálně platného principu dvojí artikulace jazyka, kdy z omezeného inventáře základních jednotek povstává složitý kombinační systém výrazového plánu jazyka, v tomto případě znakového charakteru. Různými grafickými prostředky tak byly rozlišeny obligatorní třídy kořenů (kvantitativně dominující) a fakultativní soustavy afixů (funkčně rozlišené na slovotvorné a gramatické). Odlišujícím prvkem jednotlivých popisů bylo rozhodnutí pro jeden ze tří možných způsobů prezentace morfematicky členěného materiálu. Abecedního (SOKOLOVÁ – MOŠKO – ŠIMON – BENKO 1999, TICHONOV 1996, POLJUHA 1983), vyhovujícího nejlépe studiu prefixace, retrográdního (RMSČ), odkrývajícího strukturu sufixální nebo slovotvorného, resp. podle kořenů (Šiška 2005, SOKOLOVÁ – OLOŠTIAK – IVANOVÁ 2005, JADACKA ET AL. 2001–2004, WORTH – KOZAK – JOHNSON 1970, KUZNECOVA



– JEFREMOVA 1986, KARPILOVS'KA 2002, AUGST 1975), který nejlépe odhaluje kombinatorický potenciál radixů. Samostatně stojí jediný elektronický slovník, resp. databáze (Polikarpov 2000), u něž je pochopitelně možný jakýkoli z uvedených přístupů a kromě toho mnoho dalších.

Limitujícím faktorem knižního vydání je také nutnost omezit rozsah analyzovaného lexika a až na drobné výjimky pracovat pouze s lemmaty, nikoli také slovními tvary. V některých případech nebyly zahrnuty internacionalismy, cizí slova, případně periferní složky lexikálního systému (archaismy, silně expresivní výrazy apod.). Propriální sféra lexikonu byla předmětem analýzy jen zřídka, nejčastěji jen takové jednotky, které jsou v motivačním vztahu k apelativům. Ač to čistě z hlediska morfematické analýzy nepředstavuje nijak zásadní problém (na segmentaci to nemá vliv), často byly restringovány gramaticko-lexikální komponenty analytických tvarů (zejm. se, si) a zjednodušovány homonymní a polysémní vztahy mezi lexémy, příp. jejich částmi (důsažné jsou jen ty distinkce, které se promítají do odlišného členění v „hloubkové“ struktuře).

Predikce budoucího vývoje souvisejí s obecným příklonem lingvistiky k moderním způsobům zpracování a zpřístupňování dat (automatizované systémy, lexikální databáze, jazykové korpusy atd.). Morfematická a slovotvorná analýza tak nutně musí následovat příklad morfologické analýzy, která se plně algoritmizovala, a proto uplatnila v četných aplikacích (vyhledávací techniky, strojový překlad, pravopisné a gramatické korektory apod.).

## 6 Metoda a metodika komplexní morf(emat)ické analýzy

### 6.0

Ať již morfematically, zabývající se systémem potenci jazyka, nebo morfická<sup>70</sup> analýza, orientovaná na skutečné užití jazyka, znamenají obecně „lineární rozklad slova na morfémy“ (ESČ), „zkoumání vnitřní struktury slov“ (SLAVÍČKOVÁ 1962: 96) nebo „[a]nalýz[u] slovního tvaru na menší nebo větší počet minimálních funkčně významových elementů“ (MČ1: 177). Předpokladem proveditelnosti je systémovost jazykových entit. Jejich komplexní povaha se mj. projevuje opakováním (paradigmatika) a kombinací (syntagmatika), což umožňuje porovnávání jednotlivých strukturních vzorců a modelů. Od analýzy slovotvorné se liší především svou stacionaritou, důrazem na výsledek. Ten je na rozdíl od závěrů procesuálně zaměřené slovotvorby objektivnější v tom, že se nemusí zabývat hypotézami směru vývoje (srov. práce > pracovat × rvát se > rvačka). Podle M. Dokulila jde o „lineární projekc[i] dvojí postupné binární analýzy na bezprostřední složky“ (bázi a formant), slovotvorné a tvarotvorné (MČ1: 177). Tomu odpovídá široce sdílené přesvědčení, že aktuální derivační šev koresponduje s hranicí mezi morfémy (viz např. SOKOLOVÁ 1999: 52).

Morfematically analýza je dnes uskutečňována třemi způsoby: 1) ručně, 2) automaticky a 3) kombinovanými metodami. První a zčásti i třetí postup dal vzniknout většině tištěných morfematically a derivačních slovníků představených v předcházející kapitole. Tento přístup se většinou vyznačuje vysokou kvalitou a podrobností zpracování, nevýhodou naopak bývá omezený rozsah analyzované slovní zásoby. Automatické metody charakterizuje pravý protiklad, tedy ve srovnání s lingvisty výrazně nižší úroveň a spolehlivost výsledků kompenzovaná prakticky neomezeným objemem zpracovávaných dat. Není nijak překvapivé, že prvním způsobem postupují lingvisté, druhým (a třetím) zase počítačové algoritmy navržené programátory. Zajímavější je empirický poznatek, že mezi matematiky a programátory je více jedinců se zájmem o lingvistiku než naopak. Postupem času se tak velmi zvolna zvyšuje lingvistická přijatelnost výsledků automatických analýz, zatímco rozsah kredibilních dat připravených jazykovědci stagnuje. A to přesto, že obecné zásady, na nichž jsou konkrétní analýzy postaveny, se v mnohém shodují.

Pokusím se tedy stručně ozřejmit základní principy obou antagonistických metod a zhodnotit je z hlediska vlastního návrhu počítačové reprezentace morfematically, který má, jak vyplývá z dalšího výkladu, nejbližší ke způsobu třetímu, ačkoli zcela v jiné podobě, než jaká dnes převládá. Již několikrát jsem se v předchozích kapitolách snažil poukázat na oscilace poznávání motivovanosti mezi jednotkami slovní zásoby, čemuž nasvědčují i některé připomínané poznatky z empirických výzkumů kognitivních věd. Jazykový systém je naopak abstrakce neodpovídající ani sjednocení povědomí všech badatelů, kteří se morfematically výzkumem konkrétního jazyka kdy zabývali. Jazykový popis morfematically struktury by tak jistě měl být co nejkonzistentnější a pokud možno objektivní. To však neznamená, že by měl být konečný a neměnný. Čistě sémioticky by i sebemenší proměna jediného prvku vyvolává v té či oné míře změnu celého systému. Taková otevřenost vůči konsekvantním inovacím je neuskutečnitelná prostředky jakékoli statické deskripce (žádný tištěný slovník např. nezpřístupní vztahy typu M:N, proto i každý ambiciózní pokus o zpracování morfematically dimenze jazyka dnes nemůže volit jinou než elektronickou formu zpracování. E. Slavičková si

<sup>70</sup> Termínu morfická analýza, který reflektuje distinkci *morfém* (langue) a *morf* (parole), se v praxi téměř nevyužívá, což je jednak důsledek většího lingvistického zájmu o systém (viz předchozí kapitola), zároveň však skutečnost vyvolaná terminologickým přetížením výrazu *morf*. Jak dokládá obsah první i čtvrté kapitoly, *morf* může být také chápán jako forma, vehikulum morfému, jemuž odpovídá v rovině významové sémém nebo gramém. Protože se však někdy vyjadřují různé aspekty problematiky, koexistují vedle sebe (srov. SOKOLOVÁ 1999) *morfém* (např. radix) a *submorf* (např. morf tématu). Morfematically analýza je tak termín bezpříznakový, užívaný pro všechny možné intence použití (srov. SLAVÍČKOVÁ 1962 a 1967).

toho byla již před čtyřiceti lety velice dobře vědoma, některé morfematické práce z nedávné doby (např. SOKOLOVÁ – MOŠKO – ŠIMON – BENKO 1999 nebo ŠIŠKA 2005) však počítač využívají jen jako nástroj pohodlné editace (oficiální elektronické verze neexistují, výzkum je ukončen okamžikem knižního vydání). Není tedy možné nikomu zazlívat snahy o vylepšování výsledků automatických algoritmů v řádu desetin procenta za současného tušení, že horní limit úspěšnosti je pod hranicí lingvistické přijatelnosti, pokud (zatím) žádná alternativa není.

## 6.1 Metody ruční morfematické analýzy

Předchozí řádky mohou snadno vyvolávat očekávání, že počet analytických metod je podobně jako např. definic základních pojmů značný, ve skutečnosti jde o jediný princip doplňovaný o různé množství upřesňujících podmínek, resp. podnětů, které však nikdy nebyly a ani nemohly být pro výchozí analýzu v plné míře využity. Základním kritériem všech ručních metod je sledování syntagmatické i paradigmatické **opakovatelnosti** jednotek až na minimální hranici. Odlišujícím prvkem se pak stává rozsah, v jakém se uplatňuje (viz kap. 5) na jednotlivé vrstvy a dílčí oblasti slovní zásoby (např. cizí slova, propria ad.), a míra, v níž se přihlíží k faktorům jiným, zejména sémantickým. Méně často (BOSÁK – BUZÁSSYOVÁ 1985) než v případě prací o automatických přístupech se objevuje odkaz na tzv. metodu čtverce, spojovanou se jménem J. H. Greenberga (1957). Její podstatou je, že hranice mezi morfémy  $M_1$  a  $M_2$  je prohlášena za prokázanou, pokud se morfémy vyskytují i v jiných kontextech: např.  $M_1$ - $M_2$  (*sněž-i(-t)*),  $M_1$ - $M_3$  (*sněž-n(-ý)*),  $M_4$ - $M_2$  (*nos-i(-t)*) a  $M_4$ - $M_3$  (*nos-n(-ý)*). Připouští se, že jeden vrchol čtverce může být i nulový (např. *na-pál(-i-t)*, *na-říd(-i-t)*,  $\emptyset$ -*pál(-i-t)* a  $\emptyset$ -*říd(-i-t)*). O výrazové variantnosti (alomorfii) nebo naopak o možné homografii se v této souvislosti nemluví, což je jedním z kritických bodů počítačového uplatnění (viz 6.2).

Lingvistické zásady segmentace naproti tomu specifikují, jaká další kritéria, resp. výjimky připadají v úvahu pro různé problematické jevy (zejm. asymetrický dualismus, perintegraci a deetymologizaci<sup>71</sup>). Nevyhnutelnou podmínkou všech přístupů<sup>72</sup> je **sémantická souvislost** porovnávaných morfémů. Nelze tak např. hledat argumenty pro vyčlenění morfému *kýl-* ve slově *kýl-n(i)* srovnáním se substantivem *kýl-a* nebo vytvářet u tvaru *žerte* smyšlenou analogii mezi *žert* a *žrát* (správně *žert-e* a *žer-Ø-te*). Výsledkem vhodné konfrontace jsou a) jednoznačná, b) potenciální nebo c) formální segmentace, jejichž relevanci slovenská morfematika **ověřuje v odborných pracích a** synchronních i diachronních (etymologických) **slovnících** na základě důmyslně **diferencované soustavy postupů** rozvíjejících metodu čtverce. Kvůli jednoznačné vydělitelnosti je navíc doporučována priorita vyšší abstrakce (od gramatických přes modifikační, derivační morfémy až k radixům).

Sokolová (1999) tak pro většinu derivátů předpokládá tři ověřovací kroky ( $A_1 + R_n$ ,  $A_n + R_1$ ,  $R_1$  (+ GM), kde  $A$  = afix,  $R$  = radix, GM = gramatický morfém), např. (*uč-i-tel-Ø*, *vel-i-tel-Ø*, *na-uč-i-t*, *na-uk-a*, *uč-i-t*), pro kompozita dokonce čtyři ( $R_1 + R_n$  a  $R_1$ ,  $R_n + R_2$  a  $R_2$ ). Dvěma kroky ( $A_1 + K_n$ ,  $A_n + K_1$ ,  $*K_1$ ) je možné potvrdit správnost segmentace u vázaných kořenů, jinými dvěma unikátních derivačních morfémů ad. Jedním postupem potom systemizační funkce afixů ( $A_1 + K_n$ ,  $*A_n + K_1$ ,  $*K_1$ ), např. *kolibř-ik-Ø*, *galen-it-Ø*, podobně u neohebných slov (a, při, na) apod.

Slavíčková oproti tomu kritérium opakovatelnosti v některých případech častěji nahrazuje **sémantickým zdůvodněním** morfematického členění ((pas)-tuch(a) od pás-t, (sek)-yr(a) od sekát ad.). To je také případ reziduálních kořenů vymezených **na základě analogie** s celým pojmenovacím a tvarotvorným typem (*mal-in(a)* podle *jeřab-in(a)*, *ostruž-in(a)* aj.). Někdy však segmentační zásady sledují jen princip prakticistní, jako v případě *morfémového uzlu*,

<sup>71</sup> Ve vymezení(ch) připomínaných v kap. 4.3.

<sup>72</sup> Platí to obecně, ale já se v dalším výkladu budu držet jen linie české (SLAVÍČKOVÁ 1962, 1965, 1967, RMSČ 1975 a zčásti ŠIŠKA 2005) a slovenské (HORECKÝ 1964, BOSÁK – BUZÁSSYOVÁ 1985, SOKOLOVÁ 1994, SOKOLOVÁ – ŠIMON – MOŠKO 1996, SOKOLOVÁ 1999, SOKOLOVÁ – IVANOVÁ – DZIAKOVÁ 2003, SOKOLOVÁ 2006 a KRÁLIK 2006).

kdy se o jeden foném dělí dva sousední morfémy (*měšť'-{á-[c]-tv]-í*, *{ru-[s]-k]-ý*) a kdy je pouze výhodnější předpokládat stabilitu kořene<sup>73</sup> a variaci (alomorfii) afixu než naopak.

Cím se však více přihlíží k významové stránce, tím začínají zásady segmentace připomínat zpiše pracovní zprávu. Tak např. studie *Zásady morfematickej analýzy v morfematickom slovníku slovenčiny* (SOKOLOVÁ – ŠIMON – MOŠKO 1996) má rozsah asi devíti stran, odpovídající část úvodní studie (SOKOLOVÁ 1999) pozdějšího slovníku potom zhruba dvojnásobnou. Metajazyk v takovém případě užívá obraty jako „nesegmentujeme XY okrem“ nebo „na rozdíl od YZ sme rozdelili“, příp. „pri jasnej motivácii (...) hranicu ponechávame“ ad. Slavičková sice ve své koncepci zdůvodňovala svá stanoviska a rozhodnutí v mnohém podobně, předtím se však zabývala i jinými metodami objektivizace přístupu k analýze než je např. jen ověřování v sekundárních pramenech a posuzování lingvistických nebo praktických důsledků příslušných rozhodnutí. Věnovala se tak třeba snahám o aplikaci *teorie pravděpodobnosti*, využití konceptu *centra a periferie*, dokonce zjišťovala segmentační preference rodilých mluvčích. Právě tento jednoduchý výzkum (viz SLAVÍČKOVÁ 1967) nabízí velmi zajímavý pohled na otázku, jak se nejpříhodněji vyrovnávat s „mírou nejistoty rozhodování“ a „určitou vágností jazykového systému.

Devíza její koncepce, tedy současné respektování významového (uvedeného jako první) a výrazového hlediska, naráží při praktické aplikaci zhruba na tyto hlavní typy problémů: 1) *alomorfie*: porovnávána forma obsahově blízkých morfémů se méně či více fonematically liší (např. *knih-a*, *kniž-ní*, *kouř*, *kuř-ák*, i supletivní *vz-í-t*, *ber-e*), 2) *homonymie*: náhodná shoda je jen v rovině výrazu (*vol-it*, *vol-at*, *vol-ský*), 3) *perintegrace*: vývojové přehodnocení vnitřní struktury slova se sémantickými důsledky na úrovni I) prefixu a kořene: a) defektní, vázaný, reziduální, jen etymologicky identifikovatelný radix (např. *zá-sob-a* podle *o-sob-a*, *pů-sob-i-t* aj.), b) zneprůhlednění v důsledku alternace (např. *na-dš-ení* podle *pro-dch-nou-t*, *vý-dech*), c) slovo-tvorná analogie (např. *po-bloud-i-l-y* podle *po-trh-l-y*, *po-blázn-ě-n-y* aj.) a II) kořene a sufixu: a) tzv. rozšíření radixu (historické přípony, např. *-v-* *sta-v*, *-n-* *sta-n*, *-t-* *stat-ek* ad.), b) archaický sufix (*-zeň* v *bá-zeň*) a III) prefixu, radixu a sufixu: splývání z obou stran (*s-věd-ek* a celozákladové deriváty *svědectví*, *vysvědčení* apod.). Z těchto okruhů bylo vybráno 40 problematických slov<sup>74</sup> a osloveno 50 respondentů tří generací (mladá, střední a starší) a různého vzdělání (základní, střední, vysokoškolské), mj. také několik lingvistů. Z výsledků např. vyplynulo, že nejsnazší bylo pro většinu dotazovaných vyčlenění morfémů u sloves, nejpřesvědčivěji u radixů, bez potíží se obešla také delimitace prefixů a témat. U adjektiv se projevilo jisté kolísání na hranici mezi kořenem a sufixem, jako nejobtížnější se ukázala substantiva (prefixy byly oddělovány téměř spontánně, daleko větší nesnáze působily sufixy (i jednoduché). Nejpozoruhodnější výsledek se týkal zjištění, že laičtí respondenti měli někdy přijatelnější (protože odvážnější) výsledky než přehnaně analyticky opatrní lingvisté (např. reziduálnost kořene intuitivně přijatelná jen pro nelingvisty, u sufixů překvapivě ne).

Vyhodnocení ankety autorce poskytlo argumenty pro větší důraz na formální opakovatelnost potenciálních segmentů, protože umožňuje snáze překlenout subjektivní preference jednotlivců. Zároveň jí však podnítily k úvahám o významu centrálních (oblast jasně vymezených elementů) a periferních jevů (přechodové případy, méně zapojené, nejednoznačné) v morfematice, a to z hlediska syntagmatiky i paradigmaticky. Dospěla tak k rozlišení morfémů *jaderných* (např. *-věd-* u *věd-a*, *do-věd-ět*) a *rozšiřujících* (např. *po-v po(-)věd-ět*, *na-po(-)věd-ět*, *do-po(-)věd-ět*). Vedena cílem vydělit minimální znakové





<sup>73</sup> Někdy je však nevyhnutelné počítat i s poněkud defektní podobou radixu (např. *pa-desát-Ø* × *pat-ná-ct*)

<sup>74</sup> Byla to slova *důstojník*, *svědek*, *světlo*, *poklička*, *pošetilý*, *zásoba*, *zdvihat*, *živnost*, *pondělí*, *bahno*, *zkoumat*, *čtvrtek*, *stálost*, *nadšení*, *utíkat*, *ostatní*, *zaměstnání*, *burácet*, *rozpačitý*, *složité*, *pokuta*, *křídlo*, *úkol*, *teplo*, *lázeň*, *spára*, *napětí*, *unavit*, *dovolená*, *prádélko*, *zapomenutí*, *spravedlnost*, *důvod*, *majetek*, *smrtelný*, *vrata*, *dobytek*, *setrvačnost*, *trpělivý* a *modlitba* (Slavičková 1967).

jednotky stanovila pro posuzování morfémů – na paradigmatické ose – tato čtyři kritéria: 1) *četnost výskytu v lexiku*, 2) *míra kombinační valence*, 3) *derivační potence*, 4) *polysémie* – míra rozrůzněnosti: a) *významová samostatnost* (sáz-et, saz-enice, sáz-ka, saz-ba, b) *významová vázanost* při splývání (*po(-)věd-ět, s(-)věd(-)č-it*). Dělení na jaderné (centrální) a rozšiřující (periferní) elementy Slavičková uplatnila i na sufixy, kde však mezi takovými morfémy existuje volnější vztah než u kořene. Navíc zde zavedla distinkci *funkční a prázdné*, např. adjektivní sufixální morfémy typu -ov-, -n-, -l-, které ztratily motivaci (jaderné -ov- ve *vepř-ov-ý* × rozšiřující ve *vepř-ov-ic-e*, nebo rozšiřující -ov-n- v *šach-ov-n-ic-e*, příp. původní téma -a-, -i- v *u-kaz-a-tel, děl-í-tk-o*), nebo některá deminutiva (např. prázdné -áč- v *mil-áč-ek*).

Horizont ručního, totiž na lidském úsudku založeného zpracování Slavičková překročila nástinem využití teorie pravděpodobnosti (SLAVÍČKOVÁ 1962) jakožto způsobu, který by umožnil sledovat „libovolný aspekt gramatického systému, např. způsoby tvoření slov, problémy gramatické homonymie, prediktabilitu určitých morfémových kombinací atd.“ (ibid.: 103). Praktickým způsobem, jak tohoto cíle dosáhnout, bylo stanovení tzv. podmíněné pravděpodobnosti přechodu morfémů (vztahů morfému a jeho pozice ve slově) pro jednotlivé morfémy. Nejprve bylo vždy stanoveno frekvenční pořadí morfémů předcházejících zkoumaný morfém a frekvenční pořadí morfémů za ním následujících, takže např. pravděpodobnost výskytu analyzovaného morfému na pozici bezprostředně po kořenu K se rovná podílu počtu případů, kdy za ním skutečně následuje, a úhrnné frekvence výskytu onoho radixu. Není mi známo, že by se kromě plánu „sestavit pořadí morfémů podle jejich ‚míry kombinačních schopností‘, tj. schopnosti spojovat se s jinými morfémy“ (ibid.) Slavičková tímto směru výzkumu nějak významněji dále věnovala. Vzhledem k omezeným technickým možnostem druhé poloviny šedesátých let (20. století) by to ostatně bylo docela neobvyklé. I tak je pouhá intence projevem značně průkopnického ducha.

Ve své podstatě nejde totiž o nic jiného než o návrh – dnešním slovníkem – hybridního přístupu k morfematické analýze (výše označeného číslem 3), který explicitně předpokládá ruční stanovení hranic a spočítání pravděpodobností, implicitně potom další, tentokrát již automatické využití pro segmentaci nezahrnuté slovní zásoby. Informaticky řečeno, bylo by to tzv. semi-supervised (částečně kontrolované) řešení spadající do oblasti strojového učení. Přehled ručně uplatnitelných zásad morfematické segmentace by jistě mohl pokračovat dále a sestoupit na mnohem podrobnější úroveň, z dobrých důvodů ho však na tomto místě ukončím, i když jsem dosud popsal jen základní principy. Je nepochybné, že i ty nejpodrobnější, dosud publikované návody, jak v konkrétních případech postupovat, by musely být značně aktualizovány a rozšířeny, pokud by např. někdo chtěl analyzovat dosud opomíjené vrstvy slovní zásoby, především nepřehledně širokou oblast proprií. Pro účely srovnání s automatickými přístupy, domnívám se, zaznělo vše podstatné. Již bez dalšího komentáře zakončím tuto část převzatým souhrnem klasifikačních kritérií využitelných pro identifikaci (interpretaci) delimitovaných morfémů. Uvedená míra podrobnosti velmi signifikantně ilustruje možnosti (poměrně dlouhodobého) ručního zpracování omezené slovní zásoby.

IDENTIFIKÁCIA SEGMENTOV							
POĎĽA VÝZNAMU					POĎĽA POZÍCIE V GRAMATICKOM TVARE		
SEGMENT	+ význam	lexikálny význam	KOREŇOVÉ MORFÉMY (K)		autosémantické	centrálne	sú-čas-n-ý
		slootovorný význam	DERIVAČNÉ MORFÉMY (D)		synsémantické	neviazané	list-Ø , iba
					mutačné	viazené	za-č-a:t'
					transpozičné	viacnásobné	les-o-park-Ø
	modifikačné	samostatné	t-á ist-á				
- význam	lexikálno-gramatický význam	MODIFKAČNÉ MORFÉMY (M)		aspekt	prefixálne	vy-rob-i:t'	
	gramatický význam	GRAMATICKÉ MORFÉMY (G)		gradácia	sufixálne	mam-k-a	
				negácia	postfixálne	k-ým-si	
				formové	rozštiepené	pod-kolien-k-y	
				kongruenčné	rozložené	na-jes-t' sa	
				relačné	rozložené	ven-ov-a:t' sa	
					aspekt	ne-rob-i:š	
					gradácia	nov-š-í	
					negácia	naj-nov-š-í	
						nie som	
						rob-e:n-ý	
						rob-iac	
						rob-i:l-a	
						mam-a	
						pô-jd-e:m	
						rob-i:l-a som	
						id-e:Ø sa	

Obr. 21: Klasifikace segmentů v slovenštině (Sokolová 1999: 30)

## 6.2 Automatické (a kombinované) metody morfematické analýzy

### 6.2.0 Motto

„Research based on a corpus of contemporary texts performed by automatic procedure will always be strictly synchronic and therefore the results might be somewhat different, and even unusual.“ (Cvrček 2009)

### 6.2.1 Přehled počítačových přístupů k morfematické segmentaci

Jestliže u manuální analýzy znamenají snahy stanovit pokud možno objektivní kritéria segmentace hledání optimálního poměru mezi formálním (opakovatelnost) a obsahovým principem, automatické metody s významovou stránkou pracují jen okazionalně (SCHONE – JURAFSKY 2000, YUEN – CHAN – LAI – KWONG – T'SOU 2004), i tak jde však spíše o jistou aproximaci<sup>75</sup>. Motivací většiny přístupů proto není konkurovat lingvistickým výsledkům, ale mnohem častěji možnost získat s minimem časových i hmotných nákladů z neoznačovaného<sup>76</sup> korpusu libovolného jazyka nějaká dále využitelná data (pro morfologickou analýzu, určování neznámých slov, srovnávání jazyků a jejich charakteristik, systémy zpracování řeči, strojový překlad apod.). Jednotlivé segmentační algoritmy bývají obvykle jazykově nezávislé, na nějaké úrovni byla zpracována většina jazyků s dostatečným množstvím textů v elektronické podobě, tedy i čeština, i když zájem o jednotlivé jazyky je

<sup>75</sup> Nejčastěji je využívána technika známá jako *Latent semantic analysis*, jejímž cílem je (např. pro účely rešerší) určování vztahů mezi texty na principu sumárních významových vektorů (tématu) podle onomaziologického modelu slovní zásoby (tzv. ontologie), např. *WordNet* nebo *Rogetův Thesaurus*.

<sup>76</sup> Značkováním se tradičně rozumí dodatečná informace vnesaná do korpusu za účelem zhodnocení obsahu. Obvykle to zahrnuje vyznačení hranic jazykových jednotek, přiřazení základní podoby (lemmatu) slovním tvarům a určení slovního druhu, případně bohatší lingvistická interpretace.

značně diferencovaný. Mezi nejčastěji automaticky zkoumané jazyky patří bezpochyby angličtina, arabština, finština a snad i němčina. Existují dokonce i soutěže (např. *Morpho Challenge*<sup>77</sup> pořádána od roku 2005 technickou univerzitou v Helsinkách), v nichž se na pozadí tzv. zlatého standardu (ručně zkontrolovaných dat) porovnávají výsledky automatických morfematických segmentátorů v několika částečně kontrolovaných (semi-supervised) nebo tzv. neřízených (unsupervised) analytických úlohách. Prakticky pro všechny takovéto počítačové nástroje, o nichž je v literatuře alespoň zmínka, platí, že je jejich úspěšnost (vyhodnocovaná podle tradičních měř<sup>78</sup> *precision*, *recall*, resp. *F-measure*) hluboce pod nejlepšími hodnotami<sup>79</sup> pro automatickou morfologickou analýzu.

Z hlediska metod analýzy, které jsou nejčastěji počítačově implementovány, převládá připomínaná **metoda čtverce**, kterou zpravidla doplňují různé **statistické míry**, ať již prostá frekvence výskytu (na úrovni *typů* i *tokenů*<sup>80</sup>, tvarů i pouhých znaků a jejich „shluků“), různě modifikovaná entropie, tzv. (maximum) likelihood nebo specializované heuristiky (Minimum Description Length, WordFrame, Letter successor variety apod.). Jako pomocné kritérium se někdy využívá produktivita **hypotetických flektivních tříd** (kombinatorika různých poměrů mezi kmeny a sufixy v korpusu). Množiny kandidátů na prefixy, sufixy nebo kmeny jsou v několika případech indukovány i tzv. hybridní metodou kombinující statistiky (izolovaných tokenů i minimálního kontextu) a ručně (nebo poloautomaticky) vytvořená **pravidla** (LEE – CHA – LEE 2002, WICENTOWSKI 2004) nebo jinou reprezentaci znalostí (ZHANG – KIM 1990, KUUSIK 1996). Neřízené přístupy však převládají, i když je jejich neschopnost vyrovnat se výsledkům systémům zhodnocujícím lidský vklad všeobecně přijímána: „*Yet, despite the improvements and the possibility of overfitting, the performance of knowledge-free morpheme boundary detection is far below what knowledge-rich systems (i.e. rule-based) achieve.*“ (BORDAG 2008)

Kromě cíle v samotné morfematické analýze (THURMAIR 1984, PACHUNKE – MERTINEIT – WOTHE – SCHMIDT 1992, DÉJEAN 1998, CREUTZ – LAGUS 2002, CREUTZ 2003, CREUTZ – LINDÉN 2004, KESHAVA – PITLER 2005, BORDAG 2005, REHMAN – HUSSAIN 2005, CREUTZ – LAGUS – LINDÉN – VIRPIJOLA 2005, DASGUPTA – NG 2007, KURIMO – CREUTZ – VARJOKALLIO 2007, MONSON – CARBONELL – LAVIE – LEVIN 2007, BORDAG 2008, CVRČEK 2009, ZEMAN 2009), je nejčastější využití těchto metod spatřováno v možnostech vytvořit nebo zdokonalit automatické nástroje pro analýzu morfologickou (GOLDSMITH 2001, OFLAZET – NIRENBURG – MCSHANE 2001, SNOVER – JAROSZ – BRENT 2002, LEE – PAPINENI – ROUKOS – EMAM – HASSAN 2003, HAMMARSTRÖM 2006b, GRÁC 2006, ZEMAN 2007, HLAVÁČOVÁ – HRUŠECKÝ 2008), kontrolu pravopisu neboli tzv. *spelling checking* (HAMADOU 1986), příp. rozšířit jejich schopnost rozpoznávat nová, dosud neznámá slova a tvary, tedy tzv. *guessing* (NAKOV – ANGELOVA – HAHN 2001, URREA – HLAVÁČOVÁ 2005, RIESA – YAROWSKY 2006, PIASECKI – RADZISZEWSKI 2007, MANSOUR 2008, SUNITHA – KALYANI 2009). Další uplatnění nacházejí při zkoumání slovotvorné produktivity prefixů a sufixů (URREA 2000, MONSON 2004, URREA 2006, HAMMARSTRÖM 2006a), značkování korpusů včetně tzv. *disambiguace* (kontextové zjednodušení) (HEEMSKERK 1993, MIKHEEV 1997, LEZIUS – RAPP – WETTLER 1998, REICHEL – WEILHAMMER 2004, MANSOUR 2008), analyzování mezijazykových vztahů (RAPPOPORT – LEVENT-LEVI 2006) nebo v systémech pro

<sup>77</sup> <<http://www.cis.hut.fi/morphochallenge2010/>>

<sup>78</sup> *Precision* (přesnost) udává procentuální vyjádření poměru relevantních hodnot vzhledem k irelevantním, *recall* (úplnost) potom relevantních hodnot vzhledem k celkovému počtu, *F-measure* jejich tzv. harmonický průměr.

<sup>79</sup> Srov. např. SPOUSTOVÁ (2007), SPOUSTOVÁ – HAJIČ – VOTRUBEC – KRBEC – KVĚTOŇ (2007), SPOUSTOVÁ – HAJIČ – RAAB – SPOUSTA (2009).

<sup>80</sup> Tzv. typy jsou například všechny různé slovní formy v korpusu, tokeny (exempláře) zase všechny slovní tvary v korpusu (typy včetně opakování).

rozpoznávání řeči (MENGEL 1999, RONZHIN – KARPOV 2004, JONGTAVEESATAPORN – WUTIWIWATCHAI – FURUI 2006).

Nyní se pokusím tento obecný úvod konkretizovat na příkladech algoritmů, které byly dosud aplikovány na češtinu. Relativně okrajově (seznam 30 pseudoprefixů) byl český materiál analyzován v práci o extrahování inventářů afixů z korpusových zdrojů (Urrea 2006). Data pocházela z dřívějšího výzkumu (URREA – HLAVÁČOVÁ 2005), v němž bylo neřízeným způsobem analyzováno kolem 170 tisíc lemmat z korpusu SYN2000 majících frekvenci výskytu vyšší než pět. Pro extrakci kandidátů na prefixy bylo využito hodnot entropie, tzv. ekonomického principu (usouvztažení hodnot pro počáteční a koncové řetězce) a kombinované míry „afixálnosti“. Pozitivní je na publikovaných výsledcích zjištění, že v první stovce nejčastější prefixů se nevyskytl žádný falešný (nežádoucí) kandidát, mnohem méně povzbudivé je, že s klesajícím rankem chyb znatelně přibývá a že především není možné odlišit prefixy od předním členů kompozit (1. severo-, 2. proti-, 5. super-, 6. dvoj-, 30. vy-, 58. znovu-, 60- mikro- atd.). Na rozsáhlejších datech a s některými vylepšeními v oblasti statistických měr (tzv. dopředná, zpětná a diferenční entropie), která však někdy nežádoucím způsobem filtrují délku prefixů (jednopísmenné nejsou přípustné), byl vytvořen nástroj Affisix (HLAVÁČOVÁ – HRUŠECKÝ 2008). Jeho využití je myslitelné v rámci tzv. prefixového odhadování neznámých slov, neboť také zde není možné rozlišovat mezi skutečnými prefixy, prefixoidy a kompoziční bází (euro-, byro-, pseudo-, sebe-, deseti-, nísko- apod.).

V oblasti indukování flektivních tříd ze surových (neoznačkových) dat se při své účasti v připomínané soutěži *Morpho Challenge 2007* a *2008* zabýval také češtinou D. Zeman (2007, resp. 2009). Jeho přístup spočíval v hledání opakujícího se testovaného rozdělení slova právě na jeden kmen a jeden, nebo žádný sufix (2007), resp. navíc i prefix (2009), z něhož se odvozovala pravděpodobná morfologická paradigmata. Frekvence v korpusu nebyla zkoumána. Podrobnější výsledky úspěšnosti pro češtinu nejsou známy, stejný postup aplikovaný v rámci soutěže např. na finštinu dosahoval harmonizovaného průměru úspěšnosti pouhých 30 % (58,5 % precision, 20,5 % recall), a to navíc v rámci velmi liberálních zadávacích podmínek: „*Due to possible phonological changes caused by inflection of words, the segmenters are not even required to denote the exact position of the morpheme border in the word. Therefore, the only information that can be compared with a gold standard is the number of morphemes in the word, and the fact that two words share a morpheme with the same label on specified positions.*“ (ZEMAN 2007)

Jediným významnějším pokusem o komplexní automatickou analýzu morfemické stavby češtiny je tak dosud příspěvek V. Cvrčka (CVRČEK 2009), který se zabýval neřízenou segmentací slovních tvarů (typů) korpusu SYN2005. Využíval při tom známou metodu čtverce (s jedním potenciálně neobsazeným vrcholem) doplněnou statistickými údaji. U každého potenciálního rozdělení řetězce (slovního tvaru) počítal, jak ukazuje tabulka níže, s absolutní frekvencí pro tři hrany čtverce (A', B' a A'B', čtvrtá je jako arbitrární vynechána), jejich součtem ( $\Sigma$ ), součinem ( $\Pi$ ), a hodnotami zpětné ( $H_b$ ) a dopředné ( $H_f$ ) entropie. Na základě lexikálněstatických měr (MI-score a t-score) navíc vyloučil hypotézy hranic mezi často se vyskytujícími grafémy a zakázal jedno- a dvougrafémové prefixy. Kvůli obavám z toho, že by se procedura mohla pokoušet o další segmentaci kořenového morfu, se další omezení týkalo snížení počtu iterací (cyklů) na tři a redukce zkoumaných typů jen na sedmi- a vícepísmenné. I tak (zjevně nevhodně) navržená implementace (v jazyce Perl) vykazoval enormní časovou náročnost intendované analýzy (údajně pracovala tempem pouhých 100 slov, resp. 300 cyklů za hodinu). Nakonec tak byl vyhodnocen jen vzorek 217 iterací, z nichž „správných“ hranic bylo 188, tj. 86,64 %. Soudě podle appendixu doprovázejícího Cvrčkův článek (ibid.), jehož obsahem je morfemická segmentace 3004 nejčastějších českých substantiv, adjektiv, verb a adverbii (srov. kap. 5.1, resp. STATČ 2009: 123), je uvedená hodnota jen orientační, pokud tedy nechci přistoupit na mechanické pojetí morfu.



Podle něj by totiž o tom, co je a co není morf, rozhodoval katexochén sám korpus (přístup označovaný jako *corpus-driven*), resp. soubor tvarů v něm obsažených. Bez nějakých dalších heuristik (např. přihlížení k morfologické anotaci, resp. diambiguaci), o nichž Cvrček nemluví (od homonymie abstrahuje poukazem na velikost dat), by tak nebylo nemyslitelné vyvozování identity morfémů třeba ze srovnání *ohn(-)i* (od oheň i ohnout) a *ohn(-)out* nebo *klub-u* (od klub i klubat se) a *vy-klub-al*. Ani (jakkoli ojedinělý) přístup počítající s vnitřní změnou vokálů (WICENTOWSKI 2004) by neměl možnost rozlišit např. vztah typu *na-dch-l* – *na-dš-en*. Naštěstí se např. při morfologickém značkování korpusů volí uvážlivější metoda, kdy morfologický analyzátor dokáže (na základě manuálního modelování systému) rozpoznat celé paradigma. Uváděná míra úspěšnosti Cvrčkova nástroje (pouze precision, recall bude zjevně hodně nízký), naopak implicitně pokládá za správné hranice morfů např. *-ství*, *-oval*, *-ního*, *-ační*, *-tnou* ad. (CVRČEK 2009: Appendix).

Segmentation	A'	B'	A'B'	Σ	Π	H <sub>b</sub>	H <sub>f</sub>
<i>educat-i</i>	25905	5	12616	38526	1634087400	1.75	-0.71
<i>educa-ti</i>	5606	8	4116	9730	184594368	1.53	-0.83
<i>educ-ati</i>	2357	11	2922	5290	75758694	2.58	-0.37
<i>edu-cati</i>	245	17	230	492	957950	1.04	0.66
<i>ed-ucati</i>	1	211	45	257	9495	-10.75	2.51
<i>e-ducati</i>	1	5831	850	6682	4956350	-8.05	2.79

Tab. 1: Ukázka automatické morfematické segmentace (Cvrček 2009)

### 6.3 Univerzální metoda počítačové reprezentace morfematické analýzy

Srovnání ručního (lingvistického) a automatického (programátorského) přístupu k morfematické analýze ukazuje, že ani jeden nesplňuje požadavek univerzální použitelnosti pro jednotné zpracování celého systému slovní zásoby češtiny. Manuální proto, že prakticky opomíjí většinovou část lexika (*propria*) a že nijak nezaznamenává, podle jakých kritérií (pramenů) dospěl ke konkrétnímu segmentačnímu kroku zejm. značkových slov, automatický zase neschopností zohlednit alespoň v minimální míře sémantickou stránku a základní typy alomorfie. Pokud manuální způsoby segmentace v některých případech berou v potaz etymologické výklady (srov. SOKOLOVÁ 1999 i SLAVÍČKOVÁ 1967), proč není možné stanovit nějaké objektivní zásady, jaké časové či formální hranice nejsou překročitelné (RMSČ delimituje např. *ob-ou-s-měr-n-ý* × Morfematický slovník slovenčiny *ob-oj-smer-n-ý*)? Podobně u *proprií*: má se u apelativně-*propriálních* homonym automaticky delimitovat jen *radix* a gramatický morfém flexe (např. antroponymum *Zámečník-Ø* × *zá-meč-n(-)ík-Ø* a toponym (Rychnov nad) *Kněžn-ou* × *kněž-n-ou* nebo *Příbor-Ø* × *pří-bor-Ø*), nebo postupovat v souladu s úrovní dnešního etymologického a onomastického poznání diferencovaně (tedy *kněž-n-ou* a *Kněž-n-ou*, ale *pří-bor-Ø* a *Příbor-Ø* z 'Freiburg')?

I když teď částečně předjímám některé kruciólní rysy svého návrhu, představeného v osmé kapitole, musím kategoricky prohlásit, že žádná univerzální metoda počítačové reprezentace morfematické analýzy neexistuje. Tedy ne taková, kterou by bylo možné vtělit do nějaké teoretické koncepce či manuálu. U vědomí podnětů druhé kapitoly této práce (viz 2.5) je myslitelné uvažovat pouze o dynamickém elektronickém systému znalostí, modulů a pravidel (vnitřně propojeném), který bude jakákoli rozhodnutí či naopak rektifikace promítat přímo do všech souvztažných částí systému a udržovat informace o tom, podle jakých impulzů (pramenů) je ta či ona zásada formulována, resp. výjimka stanovena. Pro češtinu musí být bez sebemenších pochybností základem slovník E. Slavičkové, protože zkrátka nic lepšího k dispozici není. To ovšem neznamená, že se obejde bez jakýchkoliv dalších korekcí. Při vší úctě, kterou k dílu Slavičkové chovám, nemohu – pokud nejde o nepochopené sémantické nuance – zavírat oči před nekonzistencemi typu *plyn* × *plou-ti* × *ply-nou-ti*, *svět-l-*

*o-un-k-ý × tep-l-oun-k-ý, u-mr-t-v-e-n-ý × mr-tv-ič-n-ý, před-vol-e-b-n-í × vol-eb-n-í nebo min-ul-e × mi-nou-ti × u-s-trn-u-l-e?* Představu konkrétního postupu ozřejmí kap. 8.2. Na samý závěr této úvahy se nemohu zříci<sup>81</sup> opakovaného, dvě stě let starého citátu z Dobrovského, s jehož podstatou se systematicky nevyrovnává ani starší český, ani novější slovenský morfematický slovník: „*Česká podstatná jména jsou buď kořenná slova nebo slova odvozená. Všechna pouhá kořenná slova jsou jednoslabičná; avšak ne každé jednoslabičné slovo je také kořeným slovem, na př. duch, čich, zrak, mlejn, stan, smrt, stav, nejsou pouhá kořenná slova, nýbrž podstatná jména vytvořená pomocí odvozovací hlásky ch, k, n, t, v a odvozená od sloves douti (duti), číti, zříti, mléti, státi, mříti.*“ (DOBROVSKÝ 1799/1953: 291, přel. B. Jedlička)

#### 6.4 Shrnutí

Kapitola přibližuje dva antagonistické přístupy k morfematické analýze, a sice ruční (lingvistický) a automatický (programátorský), jejichž společným rysem je zřetel k syntagmatické i paradigmatické opakovatelnosti jednotek. Manuální metoda dále přihlíží k sémantické stránce potenciálních morfémů, čímž se v různé míře vypořádává s problémy vyvolanými asymetrií jazykového znaku, alomorfií (variace formální stránky) a homonymií (náhodnou shodou vehikula), a vývojovými proměnami vnitřní struktury slova i celého systému neboli perintegrací. Její výsledky jsou obvykle velmi podrobné a v souladu s jazykovědnou teorií, na druhou stranu však mívají omezený rozsah. Samočinné segmentační algoritmy jsou naopak aplikovatelné na libovolné množství jazykových dat, uvedené problematické jevy ale v naprosté většině mimo jejich možnosti. Formální kritérium je sice doplňováno různými statistickými údaji (frekvence, entropie, kombinatorika a produktivita grafémů i potenciálních morfematických segmentů), celkové výsledky však zaostávají za očekáváním lingvistů i za úspěšností jiných typů automatických systémů (zejm. tzv. počítačovou morfologií). Z hlediska vlastního návrhu formální reprezentace morfematiky je dosavadní informatický přístup nepřijatelný svou asémantičností, ruční metoda je zase akceptovatelná jen zčásti, a to zejm. pro svou neúplnost (opomíjena je téměř celá oblast proprií) a jistou vágnost (nevyjasněnost míry využití etymologického kritéria). Načrtnuta je tak představa dynamického systému (podrobněji popsáno v kapitole 8), který by měl být schopen bránit se nekonzistencím a odstínit jednotlivé kroky a jejich zdůvodnění.

<sup>81</sup> Podle veřejnoprávní spisovnosti a mínění některých lingvistů jsem zde zřejmě použil hyperkorektní obrat, jehož in spe spisovná podoba by zněla *nemůžu zříct*, ať již SSČ uvádí, co chce.

## 7 Morfematická gramatika a morfematicky orientovaná slovtvorba

### 7.1 Stručný popis použitých dat a metod

Podstatou této kapitoly je představení systému češtiny z hlediska morfematické stavby, a to ve srovnání s tou podobou autentického užití jazyka, jakou nabízejí velké české (synchronní) korpusy. Systém zde reprezentuje *Retrográdní morfematický slovník češtiny* E. Slavičkové (RMSČ 1975), jehož obsah byl zevrubně analyzován, bohatě anotován a hned v několika ohledech rozšířen. Východiskem se stalo vyznačení morfematických hranic, částečně sémantický kód uváděný u každého hesla (viz také kap. 5.2.1.1). V souladu se základními principy návrhu počítačové reprezentace morfematické analýzy a syntézy (kap. 8), bylo každé lemma obvyklým způsobem fonologicky transkribováno<sup>82</sup>, tak aby obě podoby, (orto)grafická a (orto)fonemická, zachovávaly stejné hranice jednotek (např. *um-ě-n-o-věd-a* a *umň-e-n-o-vjed-a*). Podle běžných zdrojů (zejm. SSČ, SSJČ) byla provedena slovnědruhovú anotace<sup>83</sup> a určení některých gramatických kategorií (především rod u substantiv a vid u sloves). Každé heslo bylo dále charakterizováno údajem o počtu morfémů a slabik, z nichž se skládá, a doplněno frekvencí výskytu v korpusech.

Kvůli problémům s homonymií (homografií) a kompatibilitou slovnědruhových charakteristik bylo připojení statistik provedeno poněkud násilně prostřednictvím sumárních frekvencí pro všechny možné interpretace (např. lemma *telecí* adj. sčítá i hodnoty pro subst.), naopak u polysémie (např. *klika*) se údaj započítává jen jednou. Možné zkreslení celkových výsledků je jen nepatrné, na informaci o tom, jakou část korpusu pokrývá (viz níže) slovní zásoba RMSČ, to nic nemění, protože hranice morfémů zůstávají při slovnědruhovú konverzi beze změn. Statistiky byl čerpány ze dvou největších běžně dostupných korpusů současnosti: *SYN* obsahující cca 500 milionů tokenů a *SYN2009PUB* o velikosti asi 700 milionů. Pro jednoduchost byly hodnoty z obou korpusů sečteny, pramen označovaný pracovním jako *synČNK* tak čítá přibližně 1,2 miliardy textových slov. Slovní zásoba RMSČ (měřeno na frekvencích korpusových lemmat) pokrývá asi 93 % tohoto korpusu (viz dále).

Každý delimitovaný morfém byl opatřen speciální značkou kódující informace

***o jeho pozici v rámci slova:***

**P** – prefix, **R** – radix, **I** – interfix, **S** – sufix, **F** – postfix,

***slovnědruhovú příslušnosti:***

**S** – substantivum, **A** – adjektivum, **P** – pronomen, **N** – numerale, **V** – verbum, **C** – adverbium (circumstantivum), **R** – prepozice, **K** – konjunkce, **T** – partikule, **I** – interjekce,

u kořenů ***o autosémantičnosti***, resp. ***synsémantičnosti*** (podle slovních druhů):

**A** – autosémantický, **S** – synsémantický,

o složenosti, resp. ***o pozici (a korelaci) jednotlivých radixů:***

**0** – radix u nesložených slov, **12, 13, 14, 22, 23, 24, 33, 34, 44** (1. číslice = pozice, 2. číslice = z kolika celkem),

***o druhu morfému:***

<sup>82</sup> U cizích slov bylo z ekonomických důvodů oproti ortoepickým zásadám pro domácí slovní zásobu systematicky rozlišováno jen /di, ti, ni × d'i, t'i, n'i/ a spojení vokálů včetně rozlišování diftongů (např. io /ijo/, au, eu /a\_u n. au, e\_u n. eu/, gui, qui > /gvi, kvi/ apod.)

<sup>83</sup> V případě slovníkem nereflektované homonymie byla kvůli zachování rozsahu díla vybrána jedna hodnota základní, zbývající interpretace byly evidovány jen v poznámce (např. *tak* = adv. (konj., part., interj.).

**g** – gramatický, **s** – slovotvorný, **k** – kmenotvorný, **g:s** – pův. gramatický přehodnocený v slovotvorný (např. modl-**i-t**-b-a), **k:s** – pův. kmenotvorný přehodnocený v slovotvorný (např. modl-**i**-t-ba), **s:s** – pův. slovotvorný s jinou funkcí (po konverzi z jiného slov. druhu) (např. pode-zř-**i-v**-a-v-ě),

**o jeho vázanosti** (funkci rozšiřující části kořene, nebo jaderné funkci radixu):

v – vázaný (např. na-**z**-dvih-ov-a-ti).

Z praktického hlediska je využita notace XML, morfém je tak ohraničen úvodní a ukončující značkou (kvůli úspoře místa ve zjednodušené podobě):

<[PRISF] [SAPNVCRKTI] [AS] [0][1–4][1–4]] [g|s|k|g:s|k:s|s:s] [v]>**morfém**</[PRISF]>

První dvě pozice značky jsou obligatorní, další fakultativní (tzv. kompaktní).

Původní morfematické hranice byly opraveny zejm. u zjevných chyb a v případě (kmenotvorného) suffixu -ova- > -ov-a-. Navíc byly doplněny nulové hodnoty „Ø“ gramatických (věc-Ø) a kmenotvorných suffixů (kníž-Ø-e, nés-Ø-ti, sou-dob-Ø-ý<sup>84</sup>). Takto upravená data byla dále (opět podvojně: grafematická a fonematická podoba zvlášť) extrapolována na celá flektivní (kategoriální) paradigmata. Hranice morfémů zůstaly opět zachovány, pro každou jednotlivou funkci (danou úhrnem gramatických kategorií) byla kromě (ne)finálního gramatického morfému vymezena i množina alternujících morfémů (1–3, resp. v kombinaci s prefixem (ve-/v-, s-/se- aj.)). Pohled na morfonologické alternace je tak konkretizován prizmatem paradigmatických funkcí, kdy k nim přesně dochází. Je pochopitelné, že konstrukční principy české ortografie vytvářejí disproporci mezi objemem alternačních vztahů na straně grafematické a fonematické reprezentace.

Oproti předpokladům jsem nakonec musel upustit od záměru pojmut do této kapitoly i kvantitativní charakteristiky týkající se jednotlivých tvarů, nikoli jen lemmat. Jako nepřekonatelné se ukázaly zejm. tři překážky:

- 1) změny v tagsetu (systému značek) mezi korpusy SYN a SYN2009PUB a s tím související nemožnost kategoriálně napojit bez větších ztrát tvary z korpusu na příslušná místa vygenerovaných paradigmát<sup>85</sup>
- 2) potíže s pojetím lemmatizace v pražském systému, kde zvlášť vyniká koncepční problém lemmatu jako sběrné kategorie např. pro různorodé varianty (srov. HLAVÁČOVÁ 2009)
- 3) prostorové možnosti papírové podoby této práce<sup>86</sup>, která neumožňuje v přehledném tvaru znázornit všechny potřebné údaje např. pro verbální paradigmata (rozšířená o alternující morfémy)

Kromě obecné informace o celkových proporcích tak do následujících přehledů nemohu zahrnout žádné výsledky týkající se alternací<sup>87</sup>, i když je to z celé morfematické problematiky téma nejatraktivnější.

<sup>84</sup> Pojetí kmenotvorných suffixů (konektémů, resp. (alo)konektů) v rámci verbální, adjektivní a pronominální flexe přebírám z Komárkovy koncepce (srov. KOMÁREK 2006, resp. kap. 4.).

<sup>85</sup> Nutné je také neztrácet ze zřetele, jak mezerovitý úzus korpusy vzdor své úctyhodné velikosti nabízejí. Z hlediska formální morfologie je nepřehlédnutelné, že např. celé substantivní paradigma má doloženo jen malé množství lemmat. Jak jsem např. připomínal již v úvodu, pro půlmiliardový korpus SYN je to pouhých 413 substantiv, tedy jen 0,064 %.

<sup>86</sup> Plánuji však zpřístupnit všechny tyto údaje v dynamické podobě na internetové adrese <<http://www.morfemy.cz/>>, kde byly dosud k dispozici jen velmi omezené výsledky prvních morfematických analýz.

<sup>87</sup> Slavičková sice v RMSČ uvádí seznam kořenových morfémů včetně všech variant, je to však pouhý výběr (stylisticky neutrální slovní zásoby), kromě toho se alomorfie neomezuje pouze na radixy. Navíc je bez obsáhlé

Z toho nutně vyplývá také upozornění na relativnost některých kvantitativních charakteristik. Jak jsem již podotkl dříve, jakékoli statistiky týkající se slabičného aspektu morfémů není možné z lemmat (abstrakcí jednotlivých množin tvarů) bezvýhradně přenášet na celou oblast užití jazyka, tedy v tomto případě korpusovou realitu. Počet slabik se u konkrétních členů jednoho flektivního paradigmatu dost často a dost znatelně mění. U prefixů a velké části sufixů však k žádným změnám nedochází (radixy jsou také do jisté míry sylabicky stabilní, přinejmenším u sloves, adjektiv, adverbíí a indeklinabilíí). Globální statistiky (i v případě morfémů jako takových) mají větší kredibilitu s rostoucí mírou abstrakce (nižší u konkrétní materiální podoby morfémů, vyšší u konsonanticko-vokálových vzorců a velmi vysokou u slovotvorných modelů pracujících s proměnnými prefix, radix, sufix atd., resp. preradixová, radixová a postradixová část).

Navazující kvantitativní přehledy, které zde z prostorových, nikoli ideologických důvodů nemohu pro jejich značnou rozsáhlost a pro pouhý výběr nejčastějších a nejtypičtějších hodnot interpretovat, se týkají jak paradigmatických vztahů (zastoupení různých typů, slovnědruhov a distribuce apod.), tak vztahů syntagmatických (slovotvorná kombinatorika, fonemická stavba ad.). Nesporná výhoda, kterou přináší popsaný způsob anatoce každého morfému z hlediska automatického zpracování, je vykoupěna omezeným prostorem pro ilustrativní příklady v jednotlivých tabulkách. U slova *nestejnorodý* by tak úplná podoba anotace zabrala bezmála celou šíři řádku:

<P A>ne</P><R A A 12>stej</R><S A s>n</S><I A s>o</I><R A A 22>rod</R><S A k>Ø</S><S A g>ý</S>

Ve většině případů tak spoléhám na lingvistickou představivost čtenářů. Vzhledem k naprosto marginálnímu zájmu dosavadních morfematických slovníků i teorie o propriální sféru jazyka, ponechávám prozatím i já tuto oblast stranou. Několik predikcí v této oblasti však na konci kapitoly přece jen uvádím.

## 7.2 Kvantitativní charakteristiky apelativního systému češtiny ve světle korpusové evidence

### 7.2.1 Souhrnná charakteristika systému (RMSČ) prizmatem úzu (synČNK)

#### 7.2.1.1 Podle počtu slov, slabik a morfémů

63 634	slov (lemmat) vč. homonym
62 820	slov (lemmat) bez opakování
221 173	celkový počet slabik (RMSČ)
*2 467 111 486	celkový počet slabik (synČNK)
283 883	celkový počet morfémů (RMSČ)
3 260 293 078	celkový počet morf(ém)ů (synČNK)

#### 7.2.1.2 Podle podílu apelativ a proprií

62 529	apelativ (RMSČ)
1 105	proprií (RMSČ)
1 100 851 921	apelativ (synČNK)
23 550 624	proprií (synČNK)

#### 7.2.1.2 Podle zastoupení nesložených slov a kompozit

exemplifikace nesnadné přesně zjistit, průnikem jakých paradigmat a slovních druhů uváděné řady variant vznikly. Alomorfie (poziční variace) je naopak podmnožinou takových tříd, u níž jsou všechny členy průkazně distribučně vázány. Pokud by nebylo možné tyto dvě skupiny navzájem oddělit, byly by výsledky praktických (počítačových) aplikací značně diskutabilní. Chce-li tak např. někdo vyhledat jako doplňková klíčová slova některé z odvozenin, pravděpodobně ho na dotaz „děšt“ více uspokojí expanze směrem k „pršet“ (blízké jen významově) než „dštít“, které je spjaté výrazově.

59 064	nesložená (RMSČ)
4 570	kompozita (RMSČ)
1 095 919 130	nesložená (synČNK)
28 483 415	kompozita (synČNK)

### 7.2.1.3 Podle počtu všech homonym

1 594	homonyma (RMSČ)
62 040	nehomonyma (RMSČ)
57 522 185	homonyma (synČNK)
1 066 880 361	nehomonyma (synČNK)

### 7.2.1.4 Podle množství homonym s odlišnou segmentací

51	(RMSČ)
7 639 680	(synČNK)

### 7.2.1.5 Podle podílu jednoznačné a nejednoznačné segmentace (perintegrace)

2 132	hesla s jednoznačnou segmentací (RMSČ)
61 502	hesla s jednoznačnou segmentací (RMSČ)
33 186 071	hesla s jednoznačnou segmentací (synČNK)
1 091 216 474	hesla s jednoznačnou segmentací (synČNK)

### 7.2.1.6 Podle pokrytí korpusů slovní zásobou RMSČ

#### 7.2.1.6.1 RMSČ a korpus SYN

468 829 999	RMSČ (součet frekvencí z korpusu SYN u shodných lemmat)
498 995 571	SYN [= SYN2000, SYN2005, SYN2006PUB] (součet všech frekvencí)
0,9395	RMSČ / SYN
2 339	počet lemmat RMSČ neobsažených v korpusu SYN

#### 7.2.1.6.2 RMSČ a korpus SYN2009PUB

655 572 546	RMSČ (součet frekvencí z korpusu SYN2009PUB u shodných lemmat)
708 708 931	SYN2009PUB (součet všech frekvencí)
0,925	RMSČ / SYN2009PUB
1 495	počet lemmat RMSČ neobsažených v korpusu SYN2009PUB

#### 7.2.1.6.3 RMSČ a korpus synČNK

1124402545	RMSČ (frekvence z korpusu synČNK)
1207704502	synČNK [= SYN + SYN2009PUB] (součet všech frekvencí)
0,931	RMSČ / synČNK
758	počet lemmat RMSČ neobsažených v korpusu synČNK

## 7.2.2 Struktura slovní zásoby podle číselného indexu z RMSČ

Řazeno podle zvýrazněného sloupce, celkem 56 kombinací indexů.

Číselné indexy	význam	frekv. (RMSČ)	frekv. (synČNK)	Číselné indexy	význam	frekv. (RMSČ)	frekv. (synČNK)
0	součas-autosém-domáci;	27734	429819385	0	součas-autosém-domáci;	27734	429819385
3	cizí-autosém;	11891	103055068	1	synsém;	565	313481535
0,6	součas-autosém-domáci;homonym;	10391	175833795	0,6	součas-autosém-domáci;homonym;	10391	175833795
5	expres;	2908	1655182	3	cizí-autosém;	11891	103055068
2	kompoz;	2627	10267438	4	proprium;	1539	29185152
4	proprium;	1539	29185152	1,6	synsém;homonym;	33	24617607
3,2	cizí-autosém;kompoz;	1404	7193177	3,6	cizí-autosém;homonym;	1370	12862451
3,6	cizí-autosém;homonym;	1370	12862451	2	kompoz;	2627	10267438
9	nejasn-strukt;	579	2259347	3,2	cizí-autosém;kompoz;	1404	7193177
7	eponymum;	574	1077603	4,6	proprium;homonym;	64	7019818
1	synsém;	565	313481535	9	nejasn-strukt;	579	2259347

5,6	expres;homonym;	464	814362	5	expres;	2908	1655182
3,5	cizí-autosém;expres;	357	222676	8	archaic;	344	1437624
8	archaic;	344	1437624	7	eponymum;	574	1077603
8,6	archaic;homonym;	133	138568	0,7	součas-autosém-domáci;eponymum;	1	927946
9,6	nejasn-strukt;homonym;	102	173070	4,2	proprium;kompoz;	94	881364
4,2	proprium;kompoz;	94	881364	5,6	expres;homonym;	464	814362
8,2	archaic;kompoz;	88	16031	1,2	synsém;kompoz;	69	743847
1,2	synsém;kompoz;	69	743847	0,5	součas-autosém-domáci;expres;	33	380598
4,6	proprium;homonym;	64	7019818	3,5	cizí-autosém;expres;	357	222676
7,6	eponymum;homonym;	49	73270	9,6	nejasn-strukt;homonym;	102	173070
3,8	cizí-autosém;archaic;	44	11547	8,6	archaic;homonym;	133	138568
0,5	součas-autosém-domáci;expres;	33	380598	7,6	eponymum;homonym;	49	73270
1,6	synsém;homonym;	33	24617607	1,8	synsém;archaic;	18	61458
5,2	expres;kompoz;	24	28799	6	homonym;	16	55557
7,2	eponymum;kompoz;	19	8898	6,6	homonym;homonym;	1	51538
1,8	synsém;archaic;	18	61458	5,2	expres;kompoz;	24	28799
6	homonym;	16	55557	8,2	archaic;kompoz;	88	16031
1,5	synsém;expres;	16	1771	3,2,6	cizí-autosém;kompoz;homonym;	2	11850
9,5	nejasn-strukt;expres;	15	4455	3,8	cizí-autosém;archaic;	44	11547
3,5,6	cizí-autosém;expres;homonym;	14	2545	7,2	eponymum;kompoz;	19	8898
4,5	proprium;expres;	10	8109	4,5	proprium;expres;	10	8109
7,5	eponymum;expres;	7	2548	5,9	expres;nejasn-strukt;	1	4782
8,1	archaic;synsém;	5	1489	9,5	nejasn-strukt;expres;	15	4455
2,8	kompoz;archaic;	4	216	9,8	nejasn-strukt;archaic;	3	3509
5,8	expres;archaic;	3	344	7,5	eponymum;expres;	7	2548
2,5	kompoz;expres;	3	1527	3,5,6	cizí-autosém;expres;homonym;	14	2545
9,8	nejasn-strukt;archaic;	3	3509	1,7	synsém;eponymum;	1	2216
8,5	archaic;expres;	2	59	1,5	synsém;expres;	16	1771
3,2,6	cizí-autosém;kompoz;homonym;	2	11850	2,5	kompoz;expres;	3	1527
4,8	proprium;archaic;	2	470	8,1	archaic;synsém;	5	1489
3,3	cizí-autosém;cizí-autosém;	1	549	1,3	synsém;cizí-autosém;	1	907
3,5,2	cizí-autosém;expres;kompoz;	1	7	3,3	cizí-autosém;cizí-autosém;	1	549
3,6,5	cizí-autosém;homonym;expres;	1	34	4,8	proprium;archaic;	2	470
3,2,5	cizí-autosém;kompoz;expres;	1	5	5,8	expres;archaic;	3	344
3,1	cizí-autosém;synsém;	1	105	2,8	kompoz;archaic;	4	216
7,8	eponymum;archaic;	1	0	0,4	součas-autosém-domáci;proprium;	1	207
7,5,6	eponymum;expres;homonym;	1	0	3,1	cizí-autosém;synsém;	1	105
5,9	expres;nejasn-strukt;	1	4782	2,2	kompoz;kompoz;	1	67
6,6	homonym;homonym;	1	51538	9,2	nejasn-strukt;kompoz;	1	63
2,2	kompoz;kompoz;	1	67	8,5	archaic;expres;	2	59
9,2	nejasn-strukt;kompoz;	1	63	3,6,5	cizí-autosém;homonym;expres;	1	34
0,7	součas-autosém-domáci;eponymum;	1	927946	3,5,2	cizí-autosém;expres;kompoz;	1	7
0,4	součas-autosém-domáci;proprium;	1	207	3,2,5	cizí-autosém;kompoz;expres;	1	5
1,3	synsém;cizí-autosém;	1	907	7,8	eponymum;archaic;	1	0
1,7	synsém;eponymum;	1	2216	7,5,6	eponymum;expres;homonym;	1	0
Σ		63634	1124402545			63634	1124402545

Tab. 2: Struktura slovní zásoby podle číselné indexu z RMSČ\*

## 7.2.3 Morfematická struktura češtiny podle počtu morfémů

### 7.2.3.1 Celkový přehled

Počet morfémů / lemma	frekv. (RMSČ)	frekv. (synČNK)	příklad
1	306	239280611	tradá
2	5746	249258982	dveř-e
3	10254	280689361	sněž-i-ti
4	18105	186687139	volej-bal-ist-a
5	13376	99142738	parád-i-v-Ø-ý

6	10646	51670349	papír-n-ic-tv-í-Ø
7	3763	13804916	od-kys-l-ič-ov-a-ti
8	1176	3395599	u-vol-ň-ov-a-c-Ø-í
9	227	436420	ne-po-míj-ej-í-c-n-ost-Ø
10	29	21278	z-u-ži-t-k-ov-á-v-a-ti
11	5	15150	vol-n-o-myšl-e-n-k-ář-sk-Ø-ý
12	1	2	prv-o-by-t-n-ě-po-s-pol-n-Ø-ý
Σ	63634	1124402545	

Tab. 3: Morfematická struktura češtiny podle počtu morfémů (celkový přehled)

### 7.2.3.2 Podle počtu morfémů a slovních druhů

Počet morfémů / lemma	Slovní druh	frekv. (RMSČ)	frekv. (synČNK)	Počet morfémů / lemma	Slovní druh	frekv. (RMSČ)	frekv. (synČNK)
2	A	19	20842	1	P	3	3398289
3	A	237	20245877	2	P	34	64109814
4	A	2583	50389090	3	P	49	23749986
5	A	4787	51843150	4	P	22	3779069
6	A	6163	29465427	5	P	17	152558
7	A	2393	9174417	6	P	6	8774
8	A	794	2483453	7	P	1	701
9	A	155	346438	1	R	38	133344458
10	A	21	17935	2	R	7	3356865
11	A	4	15119	3	R	3	204504
12	A	1	2	4	R	2	5913
1	C	71	25536448	6	R	4	193
2	C	332	22748675	5	R	3	33
3	C	849	24930780	2	S	5276	145037783
4	C	1052	9777710	3	S	7308	111943114
5	C	669	4235073	4	S	6706	52625320
6	C	198	1120980	5	S	4373	21347374
7	C	53	298894	6	S	2874	16005080
8	C	10	7630	7	S	1060	3457819
1	I	143	286550	8	S	329	868282
2	I	15	36683	9	S	66	81526
3	I	9	10997	10	S	7	3162
4	I	8	2146	11	S	1	31
1	K	34	74508181	1	T	16	2037431
2	K	18	1256173	3	T	2	507654
3	K	7	1290643	2	T	10	194856
4	K	1	11385	4	T	3	20966
5	K	1	292925	5	T	1	18720
1	N	1	169254	3	V	1751	94833307
2	N	34	12497288	4	V	7686	68397286
3	N	39	2972499	5	V	3505	21099175



4	N	42	1678254	6	V	1396	5020633
5	N	20	153730	7	V	256	873085
6	N	5	49262	8	V	43	36234
				9	V	6	8456
				10	V	1	181
				2	?	1	3
				<b>Σ</b>		<b>63634</b>	<b>1124402545</b>

Tab. 4: Morfematická struktura češtiny podle počtu morfémů (a podle sl. druhů)

## 7.2.4 Morfematická struktura češtiny podle počtu slabik

### 7.2.4.1 Celkový přehled

Počet slabik / lemma	frekv. (RMSČ)	frekv. (synČNK)	příklad
0	4	63503916	k
1	1632	276635338	čaj-Ø
2	9783	374778833	ouvej
3	21218	242577291	dv-oj-s-přež-í-Ø
4	20913	127557309	z-po-ne-náhl-a
5	8230	35126075	blah-o-řeč-i-ti
6	1531	3719617	ekv-i-libr-ist-ik-a
7	277	470808	za-barik-ád-ov-a-n-Ø-ý
8	44	33332	individ-u-al-iz-ac-e
9	2	26	z-desat-er-o-ná-sob-ova-ti
<b>Σ</b>	<b>63634</b>	<b>1124402545</b>	

Tab. 5: Morfematická struktura češtiny podle počtu slabik (celkový přehled)

### 7.2.4.2 Podle počtu slabik a slovních druhů

Počet slabik / lemma	slovní druh	frekv. (RMSČ)	frekv. (synČNK)	Počet slabik / lemma	slovní druh	frekv. (RMSČ)	frekv. (synČNK)
1	A	14	801356	1	P	27	67258361
2	A	1160	49695130	2	P	47	26320976
3	A	5571	64610268	3	P	43	1369652
4	A	7090	35868712	4	P	15	250202
5	A	2635	10855565	0	R	4	63503916
6	A	561	1859342	1	R	24	66776653
7	A	109	282326	2	R	19	6510633
8	A	17	29051	3	R	3	120538
1	C	77	22185850	4	R	4	176
2	C	831	41930619	5	R	3	50
3	C	1318	18462820	1	S	1367	55094838
4	C	779	4977883	2	S	7359	154060198
5	C	192	923571	3	S	10802	100135652
6	C	34	175144	4	S	6187	34562306
7	C	3	303	5	S	1830	6992079
1	I	85	234017	6	S	376	456757

2	I	46	87662	7	S	69	64838
3	I	27	11303	8	S	10	2823
4	I	13	2684	1	T	10	1060411
5	I	4	710	2	T	12	1260002
1	K	19	57104461	3	T	7	439777
2	K	23	18267276	4	T	3	19437
3	K	19	1987570	1	V	1	47732
1	N	8	6071659	2	V	247	69264948
2	N	38	7381386	3	V	3379	52253405
3	N	49	3186306	4	V	6786	51062955
4	N	36	812954	5	V	3558	16286346
5	N	8	67754	6	V	558	1228146
6	N	2	228	7	V	96	123341
				8	V	17	1458
				9	V	2	26
				2	?	1	3
				<b>Σ</b>		<b>63634</b>	<b>1124402545</b>

Tab. 6: Morfematická struktura češtiny podle počtu slabik a slovních druhů

## 7.2.5 Morfematická struktura češtiny podle slovních druhů

Slovní druhy	frekv. (RMSČ)	frekv. (synČNK)	Slovní druhy	frekv. (RMSČ)	frekv. (synČNK)	Slovní druhy	frekv. (RMSČ)	frekv. (synČNK)
A	17157	164001750	S	28000	351369491	S	28000	351369491
C	3234	88656190	A	17157	164001750	V	14644	190268357
I	175	336376	V	14644	190268357	A	17157	164001750
K	61	77359307	C	3234	88656190	R	57	136911966
N	141	17520287	I	175	336376	P	132	95199191
P	132	95199191	N	141	17520287	C	3234	88656190
R	57	136911966	P	132	95199191	K	61	77359307
S	28000	351369491	K	61	77359307	N	141	17520287
T	32	2779627	R	57	136911966	T	32	2779627
V	14644	190268357	T	32	2779627	I	175	336376
?	1	3	?	1	3	?	1	3
<b>Σ</b>	<b>63634</b>	<b>1124402545</b>	<b>Σ</b>	<b>63634</b>	<b>1124402545</b>	<b>Σ</b>	<b>63634</b>	<b>1124402545</b>

Tab. 7: Morfematická struktura češtiny podle slovních druhů

## 7.2.6 Morfematická struktura češtiny z hlediska typologie morfémů

### 7.2.6.1 Podle pozice ve slově

Typ (pozice)	frekv. (RMSČ)	frekv. (synČNK)
<F ...>	146	6014017
<I ...>	2996	12449489
<P ...>	33740	322016763
<R ...>	68321	1153186925
<S ...>	178680	1766625884
<b>Σ</b>	<b>283883</b>	<b>3260293078</b>

Tab. 8: Morfematická struktura češtiny z hlediska typologie morfémů (podle pozice ve slově)

### 7.2.6.2 Podle pozice ve slově a slovního druhu

Typ (pozice + sl. druh)	frekv. (RMSČ)	frekv. (synČNK)	Typ (pozice + sl. druh)	frekv. (RMSČ)	frekv. (synČNK)
<F C ...>	63	663968	<R A ...>	18812	170835763
<F I ...>	1	2309	<R C ...>	3546	92173911
<F K ...>	24	1343629	<R I ...>	184	344545
<F P ...>	46	3804477	<R K ...>	68	78978598
<F R ...>	1	6	<R N ...>	252	20143337
<F T ...>	11	199628	<R P ...>	187	99362036
<I A ...>	1250	4112479	<R R ...>	57	136911966
<I C ...>	200	1637316	<R S ...>	30420	361111690
<I I ...>	1	1150	<R T ...>	35	2799064
<I K ...>	2	304310	<R V ...>	14760	190526015
<I N ...>	44	1364025	<S A ...>	66630	531563626
<I P ...>	24	160268	<S C ...>	6711	85490398
<I S ...>	1432	4728053	<S g ...>	1	18209
<I T ...>	1	18543	<S I ...>	35	19069
<I V ...>	42	123345	<S K ...>	3	337088
<P A ...>	10069	79437525	<S N ...>	185	19104141
<P C ...>	1955	35052060	<S P ...>	170	112953076
<P I ...>	8	8279	<S R ...>	31	949242
<P K ...>	3	1438996	<S S ...>	68052	602700787
<P N ...>	5	1085558	<S T ...>	7	566060
<P P ...>	11	2506551	<S V ...>	36855	412924188
<P R ...>	20	2835467	<b>Σ</b>	<b>283883</b>	<b>3260293078</b>
<P S ...>	9227	102494870			
<P T ...>	5	544274			
<P V ...>	12437	96613183			

Tab. 9: Morfematická struktura češtiny z hlediska typologie morfémů (podle pozice ve slově a sl. druhu)

### 7.2.6.3 Podle celé značky s vynecháním slovního druhu

Typ (pozice + charakt.)	frekv. (RMSČ)	frekv. (synČNK)	Typ (pozice + charakt.)	frekv. (RMSČ)	frekv. (synČNK)	Typ (pozice + charakt.)	frekv. (RMSČ)	frekv. (synČNK)
<F ... g:s>	16	16036	<S ... g>	60063	810429961	<S ... g>	60063	810429961
<F ... g>	130	5997981	<R ... 0>	59023	1093966633	<R ... 0>	59023	1093966633
<I ... g:s>	884	5706086	<S ... s>	46785	346604614	<S ... s>	46785	346604614
<I ... k:s>	133	121348	<S ... k>	39528	416605932	<S ... k>	39528	416605932
<I ... s>	1979	6622055	<P ...>	32348	306378462	<P ...>	32348	306378462
<P ... s>	231	737428	<S ... g:s>	14514	109707067	<S ... g:s>	14514	109707067
<P ... v>	1161	14900873	<S ... k:s>	14473	56238784	<S ... k:s>	14473	56238784
<P ...>	32348	306378462	<R ... 12>	4455	27770257	<R ... 12>	4455	27770257
<R ... 0>	47	1692943	<R ... 22>	4424	27623608	<R ... 22>	4424	27623608
<R ... 0>	59023	1093966633	<S ... s:s>	2422	9403149	<S ... s:s>	2422	9403149
<R ... 12>	4455	27770257	<I ... s>	1979	6622055	<I ... s>	1979	6622055
<R ... 12v>	3	5205	<P ... v>	1161	14900873	<P ... v>	1161	14900873
<R ... 13>	118	527732	<I ... g:s>	884	5706086	<I ... g:s>	884	5706086

<R ... 14>	1	17568	<S ... v>	813	16965510	<S ... v>	813	16965510
<R ... 22>	4424	27623608	<P ... s>	231	737428	<P ... s>	231	737428
<R ... 23>	115	519189	<I ... k:s>	133	121348	<I ... k:s>	133	121348
<R ... 24>	1	17568	<F ... g>	130	5997981	<F ... g>	130	5997981
<R ... 33>	119	566522	<R ... 33>	119	566522	<R ... 33>	119	566522
<R ... 34>	1	17568	<R ... 13>	118	527732	<R ... 13>	118	527732
<R ... 44>	1	17568	<R ... 23>	115	519189	<R ... 23>	115	519189
<R ... v 0>	13	444564	<S ... s:g>	70	659306	<S ... s:g>	70	659306
<S ... g:s v>	1	3540	<R ... 0>	47	1692943	<R ... 0>	47	1692943
<S ... g:s>	14514	109707067	<F ... g:s>	16	16036	<F ... g:s>	16	16036
<S ... g>	60063	810429961	<R ... v 0>	13	444564	<R ... v 0>	13	444564
<S ... k:s:s>	10	7496	<S ... k:s:s>	10	7496	<S ... k:s:s>	10	7496
<S ... k:s>	14473	56238784	<R ... 12v>	3	5205	<R ... 12v>	3	5205
<S ... k>	39528	416605932	<R ... 14>	1	17568	<R ... 14>	1	17568
<S ... ks>	1	525	<R ... 24>	1	17568	<R ... 24>	1	17568
<S ... s:g>	70	659306	<R ... 34>	1	17568	<R ... 34>	1	17568
<S ... s:s>	2422	9403149	<R ... 44>	1	17568	<R ... 44>	1	17568
<S ... s>	46785	346604614	<S ... g:s v>	1	3540	<S ... g:s v>	1	3540
<S ... v>	813	16965510	<S ... ks>	1	525	<S ... ks>	1	525
<b>Σ</b>	<b>283883</b>	<b>3260293078</b>	<b>Σ</b>	<b>283883</b>	<b>3260293078</b>	<b>Σ</b>	<b>283883</b>	<b>3260293078</b>

Tab. 10: Morfemická struktura češtiny z hlediska typologie morfémů (podle celé značky s vynecháním sl. druhu)

#### 7.2.6.4 Podle celé značky

Typ (celá značka)	frekv. (RMSČ)	frekv. (synČNK)	Typ (celá značka)	frekv. (RMSČ)	frekv. (synČNK)	Typ (celá značka)	frekv. (RMSČ)	frekv. (synČNK)
<S S g>	27996	351350318	<R N A 12>	72	2659259	<P l>	8	8279
<S S s>	25724	155029649	<I A k:s>	66	67810	<R l A 12>	8	4826
<R S A 0>	25615	341268510	<R N A 22>	66	2544881	<I C k:s>	7	5176
<S V k>	18325	207645070	<S A s:g>	66	658415	<I V g:s>	7	96887
<S A g>	17158	164019496	<P C v>	63	1007159	<P C s>	7	116051
<S A k>	17137	163982310	<S N k>	60	1655435	<R l A 22>	7	4751
<S A s>	16411	154448926	<S C s v>	58	2566750	<R K S 12>	7	1619291
<R A A 0>	15544	157549462	<F C g>	57	660463	<R K S 22>	7	1619291
<S V g>	14601	190174784	<R R S 0>	57	136911966	<S R g>	7	226
<R V A 0>	14483	189962287	<R K S 0>	54	75740016	<F C g:s>	6	3505
<P V>	12188	94604902	<I S k:s>	52	46108	<R N/S A 0>	6	1647143
<P A>	9602	76120838	<S V g:s>	52	461307	<S C s:s>	6	3561
<P S>	8581	93260451	<R N A 0>	47	13061300	<S N/S g>	6	1647143
<S A k:s>	8505	22970997	<R V/A A 0>	41	45800	<S P s v>	6	2728217
<S A g:s>	6625	20594658	<I N g:s>	40	1339256	<S V/A k>	6	2109
<S S k:s>	5557	30572499	<P V/A>	40	44296	<P N>	5	1085558
<S S g:s>	4201	25932228	<R S A 13>	39	16535	<P T>	5	544274
<S S k>	3980	31300124	<F P g>	37	3791952	<I N s>	4	24769
<S C g:s>	3587	55745933	<R S A 23>	37	16481	<P V s>	4	2088
<R C A 0>	2938	85343504	<R S A 33>	37	16265	<S C s:g>	4	891
<S C s>	2698	25189041	<R A A 13>	36	238528	<S N k:s>	4	594013
<R S A 12>	2348	9722622	<R A A 23>	35	199937	<S T g:s>	4	541983

<R S A 22>	2328	9621508	<R A A 33>	35	238468	<P K>	3	1438996
<S V s>	1886	6807200	<R P A 12>	35	4156771	<R S 12v>	3	5205
<P C>	1885	33928850	<S V/A k:s>	35	43950	<R T S 12>	3	19437
<S V s:s>	1739	6944550	<R P A 22>	34	4123404	<R T S 22>	3	19437
<R A A 12>	1580	6218737	<R T S 0>	29	2760190	<S T s>	3	24077
<R A A 22>	1578	6320359	<I V s>	27	24204	<R I A 13>	2	3344
<I S s>	995	3162556	<F K g>	24	1343629	<S K s>	2	44163
<I A s>	825	2606782	<R N A 33>	23	104060	<S P g:s>	2	5177856
<P S v>	528	8846981	<S P s>	23	5020812	<S P k:s>	2	67246
<S A s:s>	477	1333274	<S R g:s>	22	943001	<S S v s>	2	22706
<I S g:s>	385	1519389	<P R>	20	2835467	<S V/A s>	2	2
<S S s v>	381	7360463	<S I g:s>	20	17176	<F I g>	1	2309
<P A v>	365	3084836	<S P k>	20	12020884	<F R g:s>	1	6
<I A g:s>	359	1437887	<R N A 13>	19	63347	<I I g:s>	1	1150
<S C k:s>	357	1973147	<R N A 23>	19	63347	<I K g:s>	1	292925
<R C A 12>	284	3110350	<S N s>	19	10291	<I K s>	1	11385
<R C A 22>	283	3111013	<I P g:s>	16	114836	<I T g:s>	1	18543
<S A s v>	250	3555025	<R C A 23>	14	203041	<P V/A v>	1	1450
<P V v>	204	1960447	<R C A 33>	14	203041	<R A A 14>	1	17568
<S S s:s>	200	1121764	<S I s>	14	1875	<R A A 24>	1	17568
<R I A 0>	166	329952	<R C A 13>	13	202962	<R A A 34>	1	17568
<I C s>	119	746927	<R S A v 0>	13	444564	<R A A 44>	1	17568
<P S s>	118	387438	<S V k:s>	13	16932	<R I A 33>	1	1672
<R V A 12>	118	258964	<F T g>	11	199628	<S A ks>	1	525
<R V A 22>	118	258964	<P P>	11	2506551	<S C g>	1	11075
<S P g>	117	87938061	<R P A 23>	10	36383	<S g v>	1	18209
<S V s v>	115	736685	<S S k:s:s>	10	7496	<S I s v>	1	18
<P A s>	102	231851	<F P g:s>	9	12525	<S K g:s>	1	292925
<S N g>	96	15197259	<R P A 13>	9	3016	<S R s v>	1	143
<R P A 0>	90	91039446	<R P A 33>	9	3016	<S R s>	1	5872
<S V/A g>	81	91599	<I P s>	8	45432	<S S g:s v>	1	3540
<I C g:s>	74	885213	<I V k:s>	8	2254	<b>Σ</b>	<b>283883</b>	<b>3260293078</b>

Tab. 11: Morfemická struktura češtiny z hlediska typologie morfémů (podle celé značky)

#### 7.2.6.5 50 nejčastějších morfémů podle formy – grafematicky (homonymně, po odstranění značky)

pořadí	morfém	frekv. (RMSČ)	frekv. (synČNK)	pořadí	morfém	frekv. (RMSČ)	frekv. (synČNK)
1.	Ø	35333	514752553	1.	Ø	35333	514752553
2.	a	19681	186449983	2.	a	19681	186449983
3.	n	16700	130952629	3.	ti	14372	179760172
4.	ti	14372	179760172	4.	ý	14191	135066719
5.	ý	14191	135066719	5.	n	16700	130952629
6.	ov	8611	35950264	6.	í	6907	86874626
7.	e	6952	66647178	7.	i	6025	75929239
8.	í	6907	86874626	8.	e	6952	66647178
9.	i	6025	75929239	9.	o	5019	57493992
10.	k	5722	48763372	10.	v	3270	52104539
11.	o	5019	57493992	11.	k	5722	48763372
12.	ě	3845	34082762	12.	ne	2031	47399978

13.	v	3270	52104539	13.	po	2996	42368591
14.	po	2996	42368591	14.	ov	8611	35950264
15.	z	2613	26812251	15.	ě	3845	34082762
16.	ost	2399	10144869	16.	na	1555	33303892
17.	vy	2395	13876787	17.	s	1859	28466899
18.	za	2181	19852470	18.	t	1363	28129396
19.	á	2045	14251067	19.	z	2613	26812251
20.	ne	2031	47399978	20.	do	885	21988534
21.	u	1965	18506860	21.	za	2181	19852470
22.	l	1884	8764938	22.	u	1965	18506860
23.	s	1859	28466899	23.	pro	1175	15863424
24.	ic	1565	12949416	24.	á	2045	14251067
25.	na	1555	33303892	25.	vy	2395	13876787
26.	roz	1498	5821611	26.	sk	1319	13160382
27.	nou	1420	7162397	27.	ic	1565	12949416
28.	t	1363	28129396	28.	c	1258	10445305
29.	sk	1319	13160382	29.	ost	2399	10144869
30.	c	1258	10445305	30.	od	1092	9635483
31.	pro	1175	15863424	31.	ek	1116	9357426
32.	pře	1137	4392447	32.	při	826	9289512
33.	č	1121	6251060	33.	l	1884	8764938
34.	ek	1116	9357426	34.	in	1111	8574910
35.	in	1111	8574910	35.	ob	602	7726914
36.	od	1092	9635483	36.	nou	1420	7162397
37.	do	885	21988534	37.	y	637	6808767
38.	tel	883	5802223	38.	č	1121	6251060
39.	při	826	9289512	39.	roz	1498	5821611
40.	it	807	4286468	40.	tel	883	5802223
41.	ec	787	5714089	41.	ec	787	5714089
42.	ík	742	5265781	42.	ík	742	5265781
43.	at	728	1703819	43.	pod	577	4398833
44.	iv	720	3422615	44.	pře	1137	4392447
45.	y	637	6808767	45.	it	807	4286468
46.	ář	627	2292036	46.	iv	720	3422615
47.	ob	602	7726914	47.	stv	507	3210355
48.	ič	591	2884878	48.	ič	591	2884878
49.	pod	577	4398833	49.	ář	627	2292036
50.	stv	507	3210355	50.	at	728	1703819
...	...	...	...	...	...	...	...
Σ	10964	283883	3260293078	Σ	10964	283883	3260293078

Tab. 12: 50 nejčastějších morfémů podle formy – grafematicky

#### 7.2.6.6 50 nejčastějších morfémů podle formy a pozice ve slově – grafematicky

pořadí	morfém /pozice	frekv. (RMSČ)	frekv. (synČNK)	pořadí	morfém /pozice	frekv. (RMSČ)	frekv. (synČNK)
1.	Ø</S>	35333	514752553	1.	Ø</S>	35333	514752553
2.	a</S>	19435	153479979	2.	ti</S>	14372	179760172
3.	n</S>	16691	130604189	3.	a</S>	19435	153479979

4.	ti</S>	14372	179760172	4.	ý</S>	14190	135027622
5.	ý</S>	14190	135027622	5.	n</S>	16691	130604189
6.	ov</S>	8611	35950264	6.	í</S>	6855	86767338
7.	í</S>	6855	86767338	7.	i</S>	5697	67048614
8.	e</S>	6717	65386063	8.	e</S>	6717	65386063
9.	i</S>	5697	67048614	9.	ne</P>	2030	46748826
10.	k</S>	5697	41927634	10.	k</S>	5697	41927634
11.	ě</S>	3801	33539916	11.	bý</R>	38	41826036
12.	po</P>	2994	38042622	12.	po</P>	2994	38042622
13.	v</S>	2786	12575786	13.	ov</S>	8611	35950264
14.	z</P>	2606	14960146	14.	v</R>	17	35885781
15.	ost</S>	2397	10144783	15.	ě</S>	3801	33539916
16.	vy</P>	2393	13873278	16.	a</R>	2	31371204
17.	za</P>	2179	14748443	17.	o</S>	1368	29351749
18.	á</S>	2044	14244804	18.	se</R>	27	28588620
19.	ne</P>	2030	46748826	19.	en</S>	344	25999936
20.	o</I>	1975	6913873	20.	na</R>	2	22978962
21.	l</S>	1884	8764938	21.	t</R>	79	21966447
22.	s</P>	1834	17091479	22.	s</P>	1834	17091479
23.	o</P>	1675	12919294	23.	z</P>	2606	14960146
24.	ic</S>	1565	12949416	24.	za</P>	2179	14748443
25.	na</P>	1553	10324930	25.	á</S>	2044	14244804
26.	u</P>	1544	12691831	26.	vy</P>	2393	13873278
27.	roz</P>	1456	5427317	27.	do</P>	884	13415146
28.	nou</S>	1416	7161665	28.	sk</S>	1319	13160382
29.	o</S>	1368	29351749	29.	ic</S>	1565	12949416
30.	sk</S>	1319	13160382	30.	o</P>	1675	12919294
31.	t</S>	1284	6162949	31.	u</P>	1544	12691831
32.	c</S>	1239	7099474	32.	v</S>	2786	12575786
33.	pro</P>	1173	10147684	33.	z</R>	1	11850269
34.	pře</P>	1134	4389543	34.	s</R>	10	11352616
35.	ek</S>	1116	9357426	35.	kter</R>	16	11219234
36.	č</S>	1091	4648304	36.	že</R>	4	11020555
37.	od</P>	1088	6115102	37.	na</P>	1553	10324930
38.	in</S>	965	7364120	38.	pro</P>	1173	10147684
39.	do</P>	884	13415146	39.	ost</S>	2397	10144783
40.	tel</S>	870	5779107	40.	zá</P>	478	9813843
41.	při</P>	825	7308582	41.	ek</S>	1116	9357426
42.	it</S>	807	4286468	42.	hod</R>	214	8968434
43.	ec</S>	787	5714089	43.	vý</P>	484	8915527
44.	ík</S>	742	5265781	44.	l</S>	1884	8764938
45.	at</S>	728	1703819	45.	m</R>	23	8668066
46.	iv</S>	720	3422615	46.	do</R>	1	8573388
47.	ář</S>	627	2292036	47.	ú</P>	463	8349472
48.	ý</S>	606	6759216	48.	o</R>	1	8309076
49.	ič</S>	591	2884878	49.	i</R>	1	7464957
50.	pod</P>	553	3449345	50.	in</S>	965	7364120
...	...	...	...	...	...	...	...
Σ	11188	283883	3260293078	Σ	11188	283883	3260293078

Tab. 13: 50 nejčastějších morfémů podle formy a pozice ve slově – grafematicky

### 7.2.6.7 50 nejčastějších morfémů podle formy a celé značky – grafematicky

pořadí	morfém (značka)	frekv. (RMSČ)	frekv. (synČNK)	pořadí	morfém (značka)	frekv. (RMSČ)	frekv. (synČNK)
1.	<S A k>Ø</S>	17135	163966694	1.	<S S g>Ø</S>	17000	206184844
2.	<S S g>Ø</S>	17000	206184844	2.	<S V g>ti</S>	14372	179760172
3.	<S V g>ti</S>	14372	179760172	3.	<S A k>Ø</S>	17135	163966694
4.	<S A g>ý</S>	14090	119463573	4.	<S A g>ý</S>	14090	119463573
5.	<S V k>a</S>	7663	45669036	5.	<S S g>a</S>	6735	86518396
6.	<S S g>a</S>	6735	86518396	6.	<S V k>Ø</S>	683	67403294
7.	<S A s>n</S>	4691	64126256	7.	<S A s>n</S>	4691	64126256
8.	<S A g>s>n</S>	4640	12458578	8.	<S P g>Ø</S>	14	48707847
9.	<S V k>ov</S>	3726	17381247	9.	<S V k>a</S>	7663	45669036
10.	<S S s>k</S>	3541	17034846	10.	<S A g>i</S>	3026	43877446
11.	<S A k>s>a</S>	3329	7491479	11.	<R V A 0>bý</R>	24	41259235
12.	<S S k>i</S>	3269	25654314	12.	<S V k>i</S>	3269	38676969
13.	<S V k>i</S>	3269	38676969	13.	<R R S 0>v</R>	1	35138385
14.	<S A g>i</S>	3026	43877446	14.	<P A>ne</P>	887	32431401
15.	<S S g>s>n</S>	2819	18265150	15.	<R K S 0>a</R>	1	31261081
16.	<S S s>n</S>	2762	16571716	16.	<S S g>e</S>	2296	30775573
17.	<S S s>ost</S>	2308	9509676	17.	<R P A 0>se</R>	1	28579133
18.	<S S g>e</S>	2296	30775573	18.	<S S k>i</S>	3269	25654314
19.	<S A s>ov</S>	2246	10266914	19.	<R R S 0>na</R>	1	22978962
20.	<S A k>s>e</S>	1955	7044596	20.	<S A s>k</S>	1638	22686183
21.	<S C g>s>ě</S>	1735	19738706	21.	<S S g>o</S>	1123	19933306
22.	<S A s>k</S>	1638	22686183	22.	<S C g>s>ě</S>	1735	19738706
23.	<S V k>nou</S>	1416	7161665	23.	<S S g>s>n</S>	2819	18265150
24.	<S S k>s>e</S>	1353	9874015	24.	<S V k>ov</S>	3726	17381247
25.	<S S k>s>a</S>	1341	5244968	25.	<S S s>k</S>	3541	17034846
26.	<S A k>s>ov</S>	1253	2209711	26.	<S S s>n</S>	2762	16571716
27.	<P V>vy</P>	1245	10426436	27.	<S C s>n</S>	1132	15099224
28.	<P V>za</P>	1159	8162273	28.	<R P A 0>t</R>	9	14848174
29.	<S C s>n</S>	1132	15099224	29.	<S P g>en</S>	8	14654186
30.	<S A s>sk</S>	1128	11798501	30.	<S P g>ý</S>	34	14200681
31.	<S S g>o</S>	1123	19933306	31.	<S V k>i</S>	190	13766290
32.	<S S s>ek</S>	1095	8985329	32.	<S A g>s>n</S>	4640	12458578
33.	<P V>po</P>	1080	12102834	33.	<P V>po</P>	1080	12102834
34.	<P V>z</P>	1069	6273255	34.	<P S>po</P>	880	12087577
35.	<I S s>o</I>	991	3162145	35.	<S P k>Ø</S>	20	12020884
36.	<S V s>s>v</S>	972	4656510	36.	<R R S 0>z</R>	1	11850269
37.	<S S k>s>á</S>	901	6260008	37.	<S A s>sk</S>	1128	11798501
38.	<P A>ne</P>	887	32431401	38.	<R P A 0>kter</R>	7	11189744
39.	<P S>po</P>	880	12087577	39.	<R R S 0>s</R>	1	10812193
40.	<S A s>c</S>	879	2104466	40.	<P V>vy</P>	1245	10426436
41.	<I A s>o</I>	824	2606086	41.	<S V g>i</S>	227	10414612
42.	<S S s>in</S>	787	5329579	42.	<S A s>ov</S>	2246	10266914
43.	<S S s>č</S>	778	3523526	43.	<R K S 0>že</R>	1	9874483
44.	<P A>vy</P>	777	1710515	44.	<S S k>s>e</S>	1353	9874015
45.	<S A g>s>l</S>	776	4099398	45.	<S S s>ost</S>	2308	9509676



46.	<P V>na</P>	768	4651228	46.	<S C g:s>e</S>	432	9431234
47.	<S A s>ic</S>	757	6634928	47.	<P C>ne</P>	307	9237089
48.	<P A>po</P>	747	6853488	48.	<S S s>ek</S>	1095	8985329
49.	<P S>ne</P>	729	2622715	49.	<P V>u</P>	723	8654514
50.	<P V>u</P>	723	8654514	50.	<R R S 0>do</R>	1	8573388
...	...	...	...	...	...	...	...
Σ	24412	283883	3260293078	Σ	24412	283883	3260293078

Tab. 14: 50 nejčastějších morfémů podle formy a celé značky – grafematicky

#### 7.2.6.8 50 nejčastějších morfémů podle formy – fonematicky (homonymně, po odstranění značky)

pořadí	morfém	frekv. (RMSČ)	frekv. (synČNK)	pořadí	morfém	frekv. (RMSČ)	frekv. (synČNK)
1.	∅	35334	514755759	1.	∅	35334	514755759
2.	í	21098	221941626	2.	í	21098	221941626
3.	a	19672	186434960	3.	a	19672	186434960
4.	ťi	14369	179714382	4.	ťi	14369	179714382
5.	e	10792	100729712	5.	e	10792	100729712
6.	N	9805	56402784	6.	ň	7126	75150017
7.	ov	8285	34147210	7.	i	5861	74982312
8.	ň	7126	75150017	8.	o	5020	57493992
9.	i	5861	74982312	9.	N	9805	56402784
10.	k	5722	48763372	10.	v	2736	50017577
11.	o	5020	57493992	11.	k	5722	48763372
12.	po	2995	42368552	12.	ne	2031	47399978
13.	v	2736	50017577	13.	po	2995	42368552
14.	s	2550	31424413	14.	bí	55	41995930
15.	vi	2445	13949387	15.	ov	8285	34147210
16.	ost	2380	10132024	16.	na	1554	33302762
17.	za	2181	19852470	17.	s	2550	31424413
18.	á	2044	14247871	18.	se	337	30088160
19.	ne	2031	47399978	19.	eN	346	26388698
20.	u	1965	18506860	20.	t	961	24759561
21.	z	1923	23854750	21.	z	1923	23854750
22.	l	1884	8764938	22.	do	885	21988534
23.	ic	1565	12949416	23.	za	2181	19852470
24.	na	1554	33302762	24.	u	1965	18506860
25.	no_u	1420	7162397	25.	pro	1176	15864554
26.	sk	1319	13160382	26.	á	2044	14247871
27.	c	1258	10445305	27.	vi	2445	13949387
28.	pro	1176	15864554	28.	sk	1319	13160382
29.	pře	1137	4392447	29.	ic	1565	12949416
30.	ek	1122	9388777	30.	že	23	11783058
31.	č	1121	6251060	31.	kter	16	11219234
32.	in	1111	8596405	32.	c	1258	10445305
33.	t	961	24759561	33.	ost	2380	10132024

34.	do	885	21988534	34.	zá	477	9813782
35.	tel	883	5802223	35.	m	103	9562171
36.	roz	881	4433745	36.	ek	1122	9388777
37.	při	826	9289512	37.	při	826	9289512
38.	ij	819	7956836	38.	ú	468	9023063
39.	ec	788	5714149	39.	ví	509	8961525
40.	it	765	4113250	40.	l	1884	8764938
41.	od	748	4936006	41.	in	1111	8596405
42.	ík	742	5265781	42.	ij	819	7956836
43.	ros	631	1403149	43.	ot	609	7944852
44.	ář	627	2292036	44.	š	167	7605973
45.	ot	609	7944852	45.	no_u	1420	7162397
46.	at	595	1667266	46.	hod	166	6820300
47.	ič	592	2884975	47.	to	24	6819524
48.	iv	562	3077763	48.	pří	317	6676435
49.	ví	509	8961525	49.	ná	468	6655185
50.	stv	506	3210148	50.	vel	135	6508613
...	...	...	...	...	...	...	...
Σ	11625	283883	3260293078	Σ	11625	283883	3260293078

Tab. 15: 50 nejčastějších morfémů podle formy – fonematically

#### 7.2.6.9 50 nejčastějších morfémů podle formy a pozice ve slově – fonematically

pořadí	morfém /pozice	frekv. (RMSČ)	frekv. (synČNK)	pořadí	morfém /pozice	frekv. (RMSČ)	frekv. (synČNK)
1.	ø</S>	35334	514755759	1.	ø</S>	35334	514755759
2.	í</S>	21045	221795241	2.	í</S>	21045	221795241
3.	a</S>	19426	153464956	3.	ti</S>	14369	179714382
4.	ti</S>	14369	179714382	4.	a</S>	19426	153464956
5.	e</S>	10514	98925753	5.	e</S>	10514	98925753
6.	N</S>	9801	56082521	6.	ň</S>	7112	75101490
7.	ov</S>	8285	34147210	7.	í</S>	5504	66054746
8.	ň</S>	7112	75101490	8.	N</S>	9801	56082521
9.	k</S>	5697	41927634	9.	ne</P>	2030	46748826
10.	i</S>	5504	66054746	10.	bí</R>	55	41995930
11.	po</P>	2994	38042622	11.	k</S>	5697	41927634
12.	s</P>	2525	20048993	12.	po</P>	2994	38042622
13.	v</S>	2465	11402064	13.	v</R>	17	35885781
14.	vi</P>	2389	13873212	14.	ov</S>	8285	34147210
15.	ost</S>	2378	10131938	15.	a</R>	2	31371204
16.	za</P>	2179	14748443	16.	o</S>	1368	29351749
17.	á</S>	2043	14241608	17.	se</R>	27	28588620
18.	ne</P>	2030	46748826	18.	eN</S>	334	26181108
19.	o</I>	1976	6913873	19.	na</R>	2	22978962
20.	z</P>	1916	12002645	20.	t</R>	56	20625850
21.	l</S>	1884	8764938	21.	s</P>	2525	20048993
22.	o</P>	1675	12919294	22.	za</P>	2179	14748443

23.	ic</S>	1565	12949416	23.	á</S>	2043	14241608
24.	na</P>	1552	10323800	24.	vi</P>	2389	13873212
25.	u</P>	1544	12691831	25.	do</P>	884	13415146
26.	no_u</S>	1416	7161665	26.	sk</S>	1319	13160382
27.	o</S>	1368	29351749	27.	ic</S>	1565	12949416
28.	sk</S>	1319	13160382	28.	o</P>	1675	12919294
29.	c</S>	1239	7099474	29.	u</P>	1544	12691831
30.	pro</P>	1174	10148814	30.	z</P>	1916	12002645
31.	pře</P>	1134	4389543	31.	z</R>	1	11850269
32.	ek</S>	1122	9388777	32.	v</S>	2465	11402064
33.	č</S>	1091	4648304	33.	s</R>	10	11352616
34.	in</S>	965	7385615	34.	kter</R>	16	11219234
35.	t</S>	905	4133711	35.	že</R>	4	11020555
36.	do</P>	884	13415146	36.	na</P>	1552	10323800
37.	tel</S>	870	5779107	37.	pro</P>	1174	10148814
38.	roz</P>	839	4039451	38.	ost</S>	2378	10131938
39.	při</P>	825	7308582	39.	zá</P>	476	9813781
40.	ij</S>	817	7954226	40.	ek</S>	1122	9388777
41.	ec</S>	788	5714149	41.	ví</P>	484	8915527
42.	it</S>	765	4113250	42.	ú</P>	465	8801225
43.	od</P>	745	4933672	43.	l</S>	1884	8764938
44.	ík</S>	742	5265781	44.	m</R>	22	8636824
45.	ář</S>	627	2292036	45.	do</R>	1	8573388
46.	ros</P>	617	1387866	46.	o</R>	1	8309076
47.	at</S>	595	1667266	47.	ij</S>	817	7954226
48.	ič</S>	591	2884878	48.	i</R>	1	7464957
49.	iv</S>	562	3077763	49.	in</S>	965	7385615
50.	stv</S>	506	3210148	50.	při</P>	825	7308582
...	...	...	...	...	...	...	...
Σ	11866	283883	3260293078	Σ	11866	283883	3260293078

Tab. 16: 50 nejčastějších morfémů podle formy a pozice ve slově – fonematically

#### 7.2.6.10 50 nejčastějších morfémů podle formy a celé značky – fonematically

pořadí	morfém (značka)	frekv. (RMSČ)	frekv. (synČNK)	pořadí	morfém (značka)	frekv. (RMSČ)	frekv. (synČNK)
1.	<S A k>Ø</S>	17135	163966694	1.	<S S g>Ø</S>	17000	206184844
2.	<S A g>i</S>	17116	163341019	2.	<S V g>ti</S>	14369	179714382
3.	<S S g>Ø</S>	17000	206184844	3.	<S A k>Ø</S>	17135	163966694
4.	<S V g>ti</S>	14369	179714382	4.	<S A g>i</S>	17116	163341019
5.	<S V k>a</S>	7662	45667771	5.	<S S g>a</S>	6735	86518396
6.	<S S g>a</S>	6735	86518396	6.	<S V k>Ø</S>	684	67406490
7.	<S A g>s>N</S>	4631	11951239	7.	<S P g>Ø</S>	14	48707847
8.	<S V k>ov</S>	3727	17382512	8.	<S V k>a</S>	7662	45667771
9.	<S S s>k</S>	3541	17034846	9.	<R V A O>bí</R>	41	41429129
10.	<S A k>s>a</S>	3329	7491479	10.	<S V k>i</S>	3268	38636197
11.	<S S k>i</S>	3271	25658016	11.	<S A s>ň</S>	2276	35709898

12.	<S V k>i</S>	3268	38636197	12.	<R R S O>v</R>	1	35138385
13.	<S A k:s>e</S>	2654	9027711	13.	<S S g>e</S>	2577	35103457
14.	<S S g>e</S>	2577	35103457	14.	<P A>ne</P>	887	32431401
15.	<S A s>n</S>	2467	28956472	15.	<R K S O>a</R>	1	31261081
16.	<S S s>ost</S>	2299	9504923	16.	<S C g:s>e</S>	2167	29169940
17.	<S A s>ň</S>	2276	35709898	17.	<S A s>n</S>	2467	28956472
18.	<S C g:s>e</S>	2167	29169940	18.	<R P A O>se</R>	1	28579133
19.	<S A s>ov</S>	2166	9783421	19.	<S S k>i</S>	3271	25658016
20.	<S S g:s>ň</S>	1945	14332698	20.	<R R S O>na</R>	1	22978962
21.	<S S s>n</S>	1844	9521938	21.	<S A s>k</S>	1638	22686183
22.	<S S k:s>e</S>	1710	11758630	22.	<S S g>o</S>	1123	19933306
23.	<S A s>k</S>	1638	22686183	23.	<S V k>ov</S>	3727	17382512
24.	<S V k>no_u</S>	1416	7161665	24.	<S S s>k</S>	3541	17034846
25.	<S A k:s>ov</S>	1253	2209711	25.	<R P A O>t</R>	8	14848172
26.	<P V>vi</P>	1243	10426409	26.	<S P g>en</S>	8	14654186
27.	<S S s>ň</S>	1211	8298122	27.	<S S g:s>ň</S>	1945	14332698
28.	<S V k>e</S>	1192	12739006	28.	<S C s>ň</S>	1089	14244626
29.	<P V>za</P>	1159	8162273	29.	<S P g>i</S>	35	14211385
30.	<S S k:s>a</S>	1156	4171318	30.	<S V k>i</S>	190	13766290
31.	<S A s>sk</S>	1128	11798501	31.	<S V k>e</S>	1192	12739006
32.	<S S g>o</S>	1123	19933306	32.	<P V>po</P>	1080	12102834
33.	<S S s>ek</S>	1101	9016680	33.	<P S>po</P>	880	12087577
34.	<S C s>ň</S>	1089	14244626	34.	<S P k>ø</S>	20	12020884
35.	<P V>po</P>	1080	12102834	35.	<S A g:s>N</S>	4631	11951239
36.	<I A s>o</I>	991	3162145	36.	<R R S O>z</R>	1	11850269
37.	<S V s:s>v</S>	965	4655303	37.	<S A s>sk</S>	1128	11798501
38.	<S S k:s>á</S>	893	6227828	38.	<S S k:s>e</S>	1710	11758630
39.	<P A>ne</P>	887	32431401	39.	<R P A O>kter</R>	7	11189744
40.	<P S>po</P>	880	12087577	40.	<R R S O>s</R>	1	10812193
41.	<S A s>c</S>	879	2104466	41.	<P V>vi</P>	1243	10426409
42.	<P V>s</P>	854	5109573	42.	<S V g>i</S>	227	10414612
43.	<I A s>o</I>	824	2606086	43.	<R K S O>že</R>	1	9874483
44.	<S S s>in</S>	787	5334952	44.	<S A s>ov</S>	2166	9783421
45.	<P V>z</P>	778	5622932	45.	<S S s>n</S>	1844	9521938
46.	<S S s>č</S>	778	3523526	46.	<S S s>ost</S>	2299	9504923
47.	<P A>vi</P>	775	1710476	47.	<P C>ne</P>	307	9237089
48.	<S A g:s>l</S>	775	4061196	48.	<S A k:s>e</S>	2654	9027711
49.	<P V>na</P>	768	4651228	49.	<S S s>ek</S>	1101	9016680
50.	<S A s>ic</S>	757	6634928	50.	<P S>s</P>	665	8898852
...	...	...	...	...	...	...	...
Σ	25382	283883	3260293078	Σ	25382	283883	3260293078

Tab. 17: 50 nejčastějších morfémů podle formy a celé značky – fonematically

## 7.2.7 Morfematická struktura češtiny z hlediska fonemického složení (CV)

**7.2.7.1 50 nejčastějších morfémů podle složení z konsonantů a vokálů – grafematicky (forma a celá značka)**

pořadí	morfém (CV)	frekv. (RMSČ)	frekv. (synČNK)	pořadí	morfém (CV)	frekv. (RMSČ)	frekv. (synČNK)
1.	<S A k>C</S>	17135	163966694	1.	<S S g>C</S>	17000	206184844
2.	<S A g>V</S>	17133	163396693	2.	<S V g>CV</S>	14372	179760172
3.	<S S g>C</S>	17000	206184844	3.	<S A k>C</S>	17135	163966694
4.	<S V g>CV</S>	14372	179760172	4.	<S A g>V</S>	17133	163396693
5.	<S V k>V</S>	12363	114570830	5.	<S S g>V</S>	10996	145165474
6.	<S S s>VC</S>	11458	72059731	6.	<R S A O>CVC</R>	8073	135851835
7.	<S S g>V</S>	10996	145165474	7.	<S V k>V</S>	12363	114570830
8.	<R S A O>CVC</R>	8073	135851835	8.	<S A s>C</S>	7798	94277287
9.	<S S s>C</S>	7809	46810650	9.	<S S s>VC</S>	11458	72059731
10.	<S A s>C</S>	7798	94277287	10.	<R A A O>CVC</R>	5581	70616613
11.	<S A k:s>V</S>	6670	18689882	11.	<S V k>C</S>	683	67403294
12.	<S A g:s>C</S>	6581	19755297	12.	<R R S O>C</R>	4	63503916
13.	<S A s>VC</S>	6240	42372242	13.	<R S A O>CCVC</R>	4183	61882081
14.	<R A A O>CVC</R>	5581	70616613	14.	<R V A O>CVC</R>	5208	56758923
15.	<R V A O>CVC</R>	5208	56758923	15.	<P A>CV</P>	4344	53568779
16.	<P v>CV</P>	5120	47572195	16.	<R V A O>CV</R>	465	51231050
17.	<S S k:s>V</S>	4800	28067285	17.	<S P g>C</S>	14	48707847
18.	<P A>CV</P>	4344	53568779	18.	<P v>CV</P>	5120	47572195
19.	<R S A O>CCVC</R>	4183	61882081	19.	<S S s>C</S>	7809	46810650
20.	<S S g:s>C</S>	4163	25737568	20.	<S C g:s>V</S>	2912	46082411
21.	<P S>CV</P>	3875	42733819	21.	<P S>CV</P>	3875	42733819
22.	<S V k>VC</S>	3726	17381247	22.	<S A s>VC</S>	6240	42372242
23.	<S S s>VCC</S>	3421	16281105	23.	<R R S O>CV</R>	10	40982458
24.	<S S k>V</S>	3271	25655296	24.	<R K S O>V</R>	2	38726038
25.	<S C g:s>V</S>	2912	46082411	25.	<R P A O>CV</R>	11	33923407
26.	<R V A O>CCVC</R>	2802	22407754	26.	<R A A O>CCVC</R>	2764	32927986
27.	<R A A O>CCVC</R>	2764	32927986	27.	<R C A O>CVC</R>	1160	31304919
28.	<R S A O>CVCC</R>	2473	30458932	28.	<R S A O>CVCC</R>	2473	30458932
29.	<R S A O>CVCVC</R>	2222	27812396	29.	<S S k:s>V</S>	4800	28067285
30.	<P v>C</P>	1856	11501209	30.	<R S A O>CVCVC</R>	2222	27812396
31.	<P v>CCV</P>	1646	11688699	31.	<S S g:s>C</S>	4163	25737568
32.	<S C s>C</S>	1639	17687896	32.	<S S k>V</S>	3271	25655296
33.	<S A k:s>VC</S>	1484	3538131	33.	<P C>CV</P>	1000	23139807
34.	<P v>V</P>	1469	12748281	34.	<R V A O>CCVC</R>	2802	22407754
35.	<S V k>CVV</S>	1416	7161665	35.	<R P A O>C</R>	21	19992631
36.	<R A A O>CVCC</R>	1404	11776616	36.	<S P g>V</S>	91	19950536
37.	<R V A O>CVCC</R>	1383	12253358	37.	<S A g:s>C</S>	6581	19755297
38.	<P A>C</P>	1335	6187092	38.	<S P g>VC</S>	12	19279678
39.	<P A>V</P>	1194	5387987	39.	<S A k:s>V</S>	6670	18689882
40.	<P S>C</P>	1176	12980713	40.	<S C s>C</S>	1639	17687896
41.	<R C A O>CVC</R>	1160	31304919	41.	<S V k>VC</S>	3726	17381247
42.	<S A s>CC</S>	1157	11915753	42.	<S S s>VCC</S>	3421	16281105

43.	<R S A 22>CVC</R>	1109	5404512	43.	<R C A 0>CCVC</R>	565	14077421
44.	<P C>CV</P>	1000	23139807	44.	<R V A 0>C</R>	97	13606974
45.	<I S s>V</I>	991	3162145	45.	<P S>C</P>	1176	12980713
46.	<P A>CCV</P>	984	3535723	46.	<P v>V</P>	1469	12748281
47.	<S V s>C</S>	972	4656510	47.	<R V A 0>CVCC</R>	1383	12253358
48.	<P v>CVC</P>	970	2639305	48.	<S P k>C</S>	20	12020884
49.	<P S>V</P>	964	11468940	49.	<S A s>CC</S>	1157	11915753
50.	<S V s>C</S>	943	3998572	50.	<R A A 0>CVCC</R>	1404	11776616
...	...	...	...	...	...	...	...
Σ	1451	283883	3260293078	Σ	1451	283883	3260293078

Tab. 18: 50 nejčastějších morfémů podle složení z konsonantů a vokálů – grafematicky (forma a celá značka)

#### 7.2.7.2 50 nejčastějších morfémů podle složení z konsonantů a vokálů – grafematicky (forma)

pořadí	morfém (CV)	poč. slabik	frekv. (RMSČ)	frekv. (synČNK)	pořadí	morfém (CV)	poč. slabik	frekv. (RMSČ)	frekv. (synČNK)
1.	C	0	72142	885690924	1.	C	0	72142	885690924
2.	V	1	67882	692608408	2.	V	1	67882	692608408
3.	CV	1	31190	516194437	3.	CV	1	31190	516194437
4.	VC	1	28773	225578473	4.	CVC	1	27774	348148724
5.	CVC	1	27774	348148724	5.	VC	1	28773	225578473
6.	CCVC	1	12171	160971125	6.	CCVC	1	12171	160971125
7.	CVCC	1	6236	64715930	7.	CVCC	1	6236	64715930
8.	CC	0	4717	61941842	8.	CC	0	4717	61941842
9.	VCC	1	4295	29042913	9.	CCV	1	4265	52966615
10.	CCV	1	4265	52966615	10.	CVCVC	2	4006	46477862
11.	CVCVC	2	4006	46477862	11.	VCC	1	4295	29042913
12.	CCVCC	1	1691	11350898	12.	CVCV	2	878	22088075
13.	CVV	1	1639	11255567	13.	VCV	2	270	12663564
14.	CCC	1	1603	11861574	14.	CCC	1	1603	11861574
15.	CCCVC	1	1452	10176960	15.	CCVCC	1	1691	11350898
16.	CVCCVC	2	1385	9243179	16.	CVV	1	1639	11255567
17.	CVVC	1	895	5104658	17.	CCCVC	1	1452	10176960
18.	CVCV	2	878	22088075	18.	CVCCVC	2	1385	9243179
19.	CCC	0	773	6076694	19.	CCVCVC	2	667	6485857
20.	VCVC	2	708	6221254	20.	VCVC	2	708	6221254
21.	CCVCVC	2	667	6485857	21.	CCC	0	773	6076694
22.	CCVVC	1	620	4543062	22.	CVVC	1	895	5104658
23.	CCCC	1	484	2311423	23.	VCCVC	2	474	4820621
24.	VCCVC	2	474	4820621	24.	CCVVC	1	620	4543062
25.	CVCCC	1	454	2884957	25.	CCVCV	2	194	3685573
26.	CVCVCC	2	441	2545814	26.	CVCC	2	260	3183525
27.	CVCVCVC	3	373	1685111	27.	CVCCC	1	454	2884957
28.	VVC	1	337	1372811	28.	VV	1	293	2708015
29.	VV	1	293	2708015	29.	CVCVCC	2	441	2545814
30.	VCV	2	270	12663564	30.	CVCCV	2	168	2482702
31.	CVCC	2	260	3183525	31.	CCCC	1	484	2311423
32.	CCVCV	2	194	3685573	32.	VCCVCC	2	151	1728863

33.	CVCCV	2	168	2482702	33.	CVCVCVC	3	373	1685111
34.	CCCVCC	1	157	824416	34.	CVCVVC	3	83	1648589
35.	VCCVCC	2	151	1728863	35.	CVCVCCVC	3	67	1622489
36.	CCVCCVC	2	142	1591207	36.	CCVCCVC	2	142	1591207
37.	CVVCC	1	137	518553	37.	CVCCVC	2	110	1478518
38.	CVCCVCC	2	131	1152704	38.	VVC	1	337	1372811
39.	VVC	2	130	597271	39.	CVCCVCC	2	131	1152704
40.	CCVCCC	1	126	163966	40.	CVCCVCVC	3	107	1059330
41.	CVVC	2	126	945425	41.	VVCVC	2	123	1008672
42.	VVCVC	2	123	1008672	42.	CVVC	2	126	945425
43.	VCCC	1	116	478562	43.	CCCVCC	1	157	824416
44.	CVCCVC	2	110	1478518	44.	CCCC	1	63	805233
45.	VCCV	2	110	217248	45.	VCCVC	2	74	762399
46.	VVCVC	3	109	438667	46.	CCVC	2	45	720830
47.	CVCCVCVC	3	107	1059330	47.	CVCVVC	2	77	659965
48.	VCCVCVC	3	101	420184	48.	VVC	2	130	597271
49.	CVV	2	92	291430	49.	VVCV	2	59	596026
50.	CCVCVCC	2	86	542395	50.	CVCVCCC	3	11	579928
...	...	...	...	...	...	...	...	...	...
Σ	299		283883	3260293078	Σ	299		283883	3260293078

Tab. 19: 50 nejčastějších morfémů podle složení z konsonantů a vokálů – grafematicky (forma)

#### 7.2.7.3 Morfém z hlediska počtu slabik

morfém (poč. slabik)	frekv. (RMSČ)	frekv. (synČNK)
0	77662	953830175
1	192965	2158142235
2	11641	136157609
3	1540	12002143
4	69	156976
5	6	3940
Σ	283883	3260293078

Tab. 20: Morfém z hlediska počtu slabik

#### 7.2.7.4 50 nejčastějších morfémů podle složení z konsonantů a vokálů – fonematicky (forma a celá značka, C – kons., S – slabikotv. kons., V – jednoduch. vok., D – diftong)

pořadí	morfém (značka/CSVD)	frekv. (RMSČ)	frekv. (synČNK)	pořadí	morfém (značka/CSVD)	frekv. (RMSČ)	frekv. (synČNK)
1.	<S A k>Ø</S>	17135	163966694	1.	<S S g>Ø</S>	17000	206184844
2.	<S A g>V</S>	17133	163396693	2.	<S V g>CV</S>	14369	179714382
3.	<S S g>Ø</S>	17000	206184844	3.	<S A k>Ø</S>	17135	163966694
4.	<S V g>CV</S>	14369	179714382	4.	<S A g>V</S>	17133	163396693
5.	<S V k>V</S>	12360	114525597	5.	<S S g>V</S>	10993	145165257
6.	<S S s>VC</S>	12164	78985537	6.	<R S A O>CVC</R>	8015	132471976
7.	<S S g>V</S>	10993	145165257	7.	<S V k>V</S>	12360	114525597
8.	<S S s>C</S>	8112	48289809	8.	<S A s>C</S>	7803	94805432
9.	<R S A O>CVC</R>	8015	132471976	9.	<S S s>VC</S>	12164	78985537
10.	<S A s>C</S>	7803	94805432	10.	<R A A O>CVC</R>	5540	70216671

11.	<S A k:s>V</S>	6670	18689882	11.	<S V k>Ø</S>	684	67406490
12.	<S A g:s>C</S>	6574	19217881	12.	<R S A O>CCVC</R>	4527	65768712
13.	<S A s>VC</S>	6296	43067588	13.	<R R S O>C</R>	4	63503916
14.	<R A A O>CVC</R>	5540	70216671	14.	<R V A O>CVC</R>	5308	55336567
15.	<R V A O>CVC</R>	5308	55336567	15.	<P A>CV</P>	4342	53568740
16.	<P V>CV</P>	5118	47572168	16.	<R V A O>CV</R>	454	50528156
17.	<R S A O>CCVC</R>	4527	65768712	17.	<S P g>Ø</S>	14	48707847
18.	<S S k:s>V</S>	4463	26514017	18.	<S S s>C</S>	8112	48289809
19.	<P A>CV</P>	4342	53568740	19.	<P V>CV</P>	5118	47572168
20.	<P S>CV</P>	3873	42733757	20.	<S C g:s>V</S>	2912	46082411
21.	<S S g:s>C</S>	3862	24281428	21.	<S A s>VC</S>	6296	43067588
22.	<S V k>VC</S>	3727	17382512	22.	<P S>CV</P>	3873	42733757
23.	<S S k>V</S>	3616	27208003	23.	<R R S O>CV</R>	10	40982458
24.	<S S s>VCC</S>	3413	16258059	24.	<R K S O>V</R>	2	38726038
25.	<R V A O>CCVC</R>	2965	23825489	25.	<R P A O>CV</R>	11	33923407
26.	<R A A O>CCVC</R>	2964	32654731	26.	<R A A O>CCVC</R>	2964	32654731
27.	<S C g:s>V</S>	2912	46082411	27.	<R C A O>CVC</R>	1126	30180661
28.	<R S A O>CVCVC</R>	2371	28398688	28.	<R S A O>CVCVC</R>	2371	28398688
29.	<R S A O>CVCC</R>	2256	25298363	29.	<S S k>V</S>	3616	27208003
30.	<P V>C</P>	1856	11501209	30.	<S S k:s>V</S>	4463	26514017
31.	<P V>CCV</P>	1646	11688699	31.	<R S A O>CVCC</R>	2256	25298363
32.	<S C s>C</S>	1634	17658912	32.	<S S g:s>C</S>	3862	24281428
33.	<S A k:s>VC</S>	1484	3538131	33.	<R V A O>CCVC</R>	2965	23825489
34.	<P V>V</P>	1469	12748281	34.	<P C>CV</P>	999	23138677
35.	<S V k>CD</S>	1416	7161665	35.	<R P A O>C</R>	21	19992631
36.	<P A>C</P>	1335	6187092	36.	<S P g>V</S>	91	19950536
37.	<R A A O>CVCC</R>	1286	10974741	37.	<S P g>VC</S>	12	19279678
38.	<R V A O>CVCC</R>	1264	12421200	38.	<S A g:s>C</S>	6574	19217881
39.	<P A>V</P>	1194	5387987	39.	<S A k:s>V</S>	6670	18689882
40.	<P S>C</P>	1177	12980726	40.	<S C s>C</S>	1634	17658912
41.	<S A s>CC</S>	1158	11915901	41.	<S V k>VC</S>	3727	17382512
42.	<R C A O>CVC</R>	1126	30180661	42.	<S S s>VCC</S>	3413	16258059
43.	<R S A 22>CVC</R>	1071	5385941	43.	<R C A O>CCVC</R>	614	14485615
44.	<P C>CV</P>	999	23138677	44.	<R V A O>C</R>	96	13575732
45.	<I S s>V</I>	991	3162145	45.	<P S>C</P>	1177	12980726
46.	<P A>CCV</P>	984	3535723	46.	<P V>V</P>	1469	12748281
47.	<P V>CVC</P>	972	2639332	47.	<R V A O>CVCC</R>	1264	12421200
48.	<S V s:s>C</S>	965	4655303	48.	<S P k>Ø</S>	20	12020884
49.	<P S>V</P>	964	11468940	49.	<S A s>CC</S>	1158	11915901
50.	<S V s>C</S>	948	3999778	50.	<P V>CCV</P>	1646	11688699
...	...	...	...	...	...	...	...
Σ	1568	2883883	3260293078	Σ	1568	283883	3260293078

Tab. 21: 50 nejčastějších morfémů podle složení z konsonantů a vokálů – fonematically (CSVD)



**7.2.7.5 50 nejčastějších morfémů podle složení z konsonantů a vokálů – fonematically (forma a celá značka, C – kons., V – vok.)**

pořadí	morfém (značka/CV)	frekv. (RMSČ)	frekv. (synČNK)	pořadí	morfém (značka/CV)	frekv. (RMSČ)	frekv. (synČNK)
1.	<S A k>Ø</S>	17135	163966694	1.	<S S g>Ø</S>	17000	206184844
2.	<S A g>V</S>	17133	163396693	2.	<S V g>CV</S>	14369	179714382
3.	<S S g>Ø</S>	17000	206184844	3.	<S A k>Ø</S>	17135	163966694
4.	<S V g>CV</S>	14369	179714382	4.	<S A g>V</S>	17133	163396693
5.	<S V k>V</S>	12360	114525597	5.	<S S g>V</S>	10993	145165257
6.	<S S s>VC</S>	12164	78985537	6.	<R S A O>CVC</R>	8015	132471976
7.	<S S g>V</S>	10993	145165257	7.	<S V k>V</S>	12360	114525597
8.	<S S s>C</S>	8112	48289809	8.	<S A s>C</S>	7803	94805432
9.	<R S A O>CVC</R>	8015	132471976	9.	<S S s>VC</S>	12164	78985537
10.	<S A s>C</S>	7803	94805432	10.	<R A A O>CVC</R>	5540	70216671
11.	<S A k:s>V</S>	6670	18689882	11.	<S V k>Ø</S>	684	67406490
12.	<S A g:s>C</S>	6574	19217881	12.	<R S A O>CCVC</R>	4527	65768712
13.	<S A s>VC</S>	6296	43067588	13.	<R R S O>C</R>	4	63503916
14.	<R A A O>CVC</R>	5540	70216671	14.	<R V A O>CVC</R>	5308	55336567
15.	<R V A O>CVC</R>	5308	55336567	15.	<P A>CV</P>	4342	53568740
16.	<P V>CV</P>	5118	47572168	16.	<R V A O>CV</R>	454	50528156
17.	<R S A O>CCVC</R>	4527	65768712	17.	<S P g>Ø</S>	14	48707847
18.	<S S k:s>V</S>	4463	26514017	18.	<S S s>C</S>	8112	48289809
19.	<P A>CV</P>	4342	53568740	19.	<P V>CV</P>	5118	47572168
20.	<P S>CV</P>	3873	42733757	20.	<S C g:s>V</S>	2912	46082411
21.	<S S g:s>C</S>	3862	24281428	21.	<S A s>VC</S>	6296	43067588
22.	<S V k>VC</S>	3727	17382512	22.	<P S>CV</P>	3873	42733757
23.	<S S k>V</S>	3616	27208003	23.	<R R S O>CV</R>	10	40982458
24.	<S S s>VCC</S>	3414	16258218	24.	<R K S O>V</R>	2	38726038
25.	<R V A O>CCVC</R>	2965	23825489	25.	<R P A O>CV</R>	11	33923407
26.	<R A A O>CCVC</R>	2964	32654731	26.	<R A A O>CCVC</R>	2964	32654731
27.	<S C g:s>V</S>	2912	46082411	27.	<R C A O>CVC</R>	1126	30180661
28.	<R S A O>CVCVC</R>	2371	28398688	28.	<R S A O>CVCVC</R>	2371	28398688
29.	<R S A O>CVCC</R>	2352	27189514	29.	<S S k>V</S>	3616	27208003
30.	<P V>C</P>	1856	11501209	30.	<R S A O>CVCC</R>	2352	27189514
31.	<P V>CCV</P>	1646	11688699	31.	<S S k:s>V</S>	4463	26514017
32.	<S C s>C</S>	1634	17658912	32.	<S S g:s>C</S>	3862	24281428
33.	<S A k:s>VC</S>	1484	3538131	33.	<R V A O>CCVC</R>	2965	23825489
34.	<P V>V</P>	1469	12748281	34.	<P C>CV</P>	999	23138677
35.	<S V k>CVV</S>	1416	7161665	35.	<R P A O>C</R>	21	19992631
36.	<P A>C</P>	1335	6187092	36.	<S P g>V</S>	91	19950536
37.	<R A A O>CVCC</R>	1320	11172140	37.	<S P g>VC</S>	12	19279678
38.	<R V A O>CVCC</R>	1269	12421235	38.	<S A g:s>C</S>	6574	19217881
39.	<P A>V</P>	1194	5387987	39.	<S A k:s>V</S>	6670	18689882
40.	<P S>C</P>	1177	12980726	40.	<S C s>C</S>	1634	17658912
41.	<S A s>CC</S>	1158	11915901	41.	<S V k>VC</S>	3727	17382512

42.	<R C A O>CVC</R>	1126	30180661	42.	<S S s>VCC</S>	3414	16258218
43.	<R S A 22>CVC</R>	1071	5385941	43.	<R C A O>CCVC</R>	614	14485615
44.	<P C>CV</P>	999	23138677	44.	<R V A O>C</R>	96	13575732
45.	<I S s>V</I>	991	3162145	45.	<P S>C</P>	1177	12980726
46.	<P A>CCV</P>	984	3535723	46.	<P V>V</P>	1469	12748281
47.	<P V>CVC</P>	972	2639332	47.	<R V A O>CVCC</R>	1269	12421235
48.	<S V s>C</S>	965	4655303	48.	<S P k>Ø</S>	20	12020884
49.	<P S>V</P>	964	11468940	49.	<S A s>CC</S>	1158	11915901
50.	<S V s>C</S>	948	3999778	50.	<P V>CCV</P>	1646	11688699
...	...	...	...	...	...	...	...
Σ	1405	283883	3260293078	Σ	1405	283883	3260293078

Tab. 22: 50 nejčastějších morfémů podle složení z konsonantů a vokálů – fonematicky (CV: forma, celá značka)

#### 7.2.7.6 50 nejčastějších morfémů podle složení z konsonantů a vokálů – fonematicky (forma, C – kons., V – vok.)

pořadí	morfém (CV)	poč. slabik	frekv. (RMSČ)	frekv. (synČNK)	pořadí	morfém (CV)	poč. slabik	frekv. (RMSČ)	frekv. (synČNK)
1.	V	1	67077	684882554	1.	V	1	67077	684882554
2.	C	0	36637	370043566	2.	CV	1	31116	515145621
3.	Ø	0	35334	514755759	3.	Ø	0	35334	514755759
4.	CV	1	31116	515145621	4.	C	0	36637	370043566
5.	VC	1	29232	235429611	5.	CVC	1	27633	340771452
6.	CVC	1	27633	340771452	6.	VC	1	29232	235429611
7.	CCVC	1	13080	166959813	7.	CCVC	1	13080	166959813
8.	CVCC	1	5921	61480236	8.	CC	0	4913	64652673
9.	CC	0	4913	64652673	9.	CVCC	1	5921	61480236
10.	VCC	1	4682	26993440	10.	CCV	1	4290	53088203
11.	CCV	1	4290	53088203	11.	CVCVC	2	4249	46709830
12.	CVCVC	2	4249	46709830	12.	VCC	1	4682	26993440
13.	CCVCC	1	1783	15509738	13.	CVCV	2	813	20676354
14.	CCC	1	1677	12195548	14.	CCVCC	1	1783	15509738
15.	CVV	1	1637	11255461	15.	VCV	2	262	12582977
16.	CVCCVC	2	1387	10180200	16.	CCC	1	1677	12195548
17.	CCCVC	1	1298	11925513	17.	CCCVC	1	1298	11925513
18.	CVVC	1	913	5035869	18.	CVV	1	1637	11255461
19.	CVCV	2	813	20676354	19.	CVCCVC	2	1387	10180200
20.	VCVC	2	806	6819762	20.	VCVC	2	806	6819762
21.	CCC	0	738	4112232	21.	CVVC	1	913	5035869
22.	CCVVC	2	710	4706762	22.	VCCVC	2	508	4846790
23.	CCVVC	1	649	4582910	23.	CCVVC	2	710	4706762
24.	VCCVC	2	508	4846790	24.	CCVVC	1	649	4582910
25.	CVCVCVC	3	471	3310329	25.	CCVCCVC	2	130	4196092
26.	CVCVCC	2	433	2746103	26.	CCC	0	738	4112232
27.	CCCC	1	420	1975106	27.	CCVCV	2	220	3718274
28.	VVC	1	330	1368112	28.	CVCVCVC	3	471	3310329
29.	VV	1	293	2708015	29.	CVCC	2	259	3160186
30.	CVCCC	1	274	1828057	30.	CVCVCC	2	433	2746103
31.	VCV	2	262	12582977	31.	VV	1	293	2708015
32.	CVCC	2	259	3160186	32.	CVCCV	2	161	2572534

33.	CCVCV	2	220	3718274	33.	VCCVCC	2	156	2137097
34.	CVCCV	2	161	2572534	34.	CCCC	1	420	1975106
35.	VCCVCC	2	156	2137097	35.	CVCCC	1	274	1828057
36.	CCVCCVC	2	130	4196092	36.	CVCVCCVC	3	80	1758564
37.	CVCCVCC	2	129	1023312	37.	CVCCVC	2	116	1600838
38.	VCCVCVC	3	124	459413	38.	CCCV	1	68	1464410
39.	VCVCVC	3	124	537376	39.	VVC	1	330	1368112
40.	VCVCC	2	123	687858	40.	CVCCVCVC	3	120	1091555
41.	CVCCVCVC	3	120	1091555	41.	CVCCVCC	2	129	1023312
42.	VCCV	2	119	347435	42.	CCCVCC	1	99	837125
43.	CVCCVC	2	116	1600838	43.	CCCC	1	62	807776
44.	CVVCC	1	106	542140	44.	VCCVC	2	74	795540
45.	CCCVCC	1	99	837125	45.	CVVC	2	73	793400
46.	CCVCVCC	2	89	789871	46.	CCVCVCC	2	89	789871
47.	CVCCC	2	82	442421	47.	CCVC	2	46	720942
48.	CVCVCCVC	3	80	1758564	48.	VCVCC	2	123	687858
49.	VCCC	1	80	400297	49.	CCVCVCVC	3	21	679987
50.	VCCVC	2	74	795540	50.	CVCVCVCC	3	60	639321
...	...	...	...	...	...	...	...	...	...
Σ	255		283883	3260293078	Σ	255		283883	3260293078

Tab. 23: 50 nejčastějších morfémů podle složení z konsonantů a vokálů – fonematically (CV: forma)

## 7.2.8 Morfemická struktura češtiny ze slovotvorného hlediska

### 7.2.8.1 Vzorce struktury slov podle typů morfémů (preradix. část | radix. č. | postradix. č.)

Řazení podle frekvence v RMSČ.

preradix. část	radixová část	postradix. část	frekv. (RMSČ)	frekv. (synČNK)
</P>	</R>	</S></S>	27078	95599497
	</R>	</S></S></S>	23817	82639032
	</R>	</S></S>	23343	201318117
</P>	</R>	</S></S></S></S>	21021	30315271
</P>	</R>	</S></S></S>	18498	53625708
	</R>	</S>	16758	238249152
	</R>	</S></S></S></S>	16269	32612036
</P>	</R>	</S></S></S></S></S>	6357	6644874
</P>	</R>	</S>	5622	65811445
	</R>	</S></S></S></S></S>	5067	7082010
</P></P>	</R>	</S></S></S>	2379	9954221
</P></P>	</R>	</S></S></S></S>	2271	4099518
	</R></I></R>	</S></S>	2133	2422084
	</R></I></R>	</S></S></S>	1983	2335651
</P></P>	</R>	</S></S>	1947	7960765
	</R></R>	</S>	1467	5130491
</P>	</R>	</S></S></S></S></S></S>	1410	836059
	</R></I></R>	</S>	1389	3197699
</P></P>	</R>	</S></S></S></S></S>	1233	1601094
	</R></R>	</S></S>	1188	2153814
	</R>	</S></S></S></S></S></S>	939	982036
	</R>		918	239280611

	</R></R>	</S></S></S>	756	1376217
	</R></R>	</S></S></S></S>	600	1055782
	</R></I></R>	</S></S></S></S>	594	383474
</P></P>	</R>	</S>	579	2622646
	</R></S></I></R>	</S></S></S>	342	168572
	</R></S></I></R>	</S></S>	279	115754
	</R></I></P></R>	</S></S></S>	231	222362
</P>	</R>		207	6150939
</P></P>	</R>	</S></S></S></S></S></S>	189	138536
</P>	</R></I></R>	</S></S></S>	186	444621
</P>	</R>	</S></S></S></S></S></S></S>	168	74741
	</R>	</F>	159	4476068
</P></P></P>	</R>	</S></S></S></S>	156	469202
	</R>	</S></S></S></S></S></S></S>	150	107574
	</R></S></I></R>	</S>	135	173473
	</R>	</S></F>	129	1170820
	</R></I></R>	</S></S></S></S></S>	117	53961
	</R></I></P></R>	</S>	114	226911
	</R></R>		114	382823
	</R></R>	</S></S></S></S></S>	111	97575
</P>	</R></I></R>	</S></S>	108	207113
</P></P></P>	</R>	</S></S></S></S></S>	99	59637
	</R></I></P></R>	</S></S></S></S>	93	16989
</P></P></P>	</R>	</S></S></S>	90	240693
</P>	</R></I></R>	</S>	87	593074
	</R></I></P></R>	</S></S>	84	54164
</P></P></P>	</R>	</S></S>	78	207635
	</R></S></I></R>	</S></S></S></S>	72	17151
</P>	</R></R>	</S></S></S>	72	25506
</P>	</R></I></R>	</S></S></S></S>	72	71567
	</R></I></R>		51	520331
	</R></S></I></P></R>	</S></S></S>	51	109054
</P>	</R></I></P></R>	</S></S></S></S>	51	14279
</P></P>	</R>		48	1416172
	</R></R></I></R>	</S></S>	45	121871
</P>	</R></S></I></R>	</S></S></S>	45	13377
	</R></R></R>	</S>	42	5921
</P>	</R></R>	</S></S></S></S>	39	221321
	</R></I></R></I></R>	</S></S>	36	108243
</P>	</R></R>	</S>	36	15257
</P></P>	</R>	</S></S></S></S></S></S></S>	36	7527
	</R></P></R>	</S>	33	854
	</R></R></R>		33	7281
</P>	</R></R>	</S></S></S></S></S>	33	29140
	</R></S></R>	</S>	30	1896
	</R></S></R>	</S></S></S>	30	4842
</P>	</R></I></P></R>	</S></S></S>	27	2834

</P>	</R></I></R>	</S></S></S></S></S>	27	6590
</P></P></P>	</R>	</S>	27	13990
	</R></R></R>	</S></S>	24	3104
</P>	</R></R>		24	1253280
	</R>	</S></S></S></S></S></S></S></S>	21	4226
	</R></I></R></I></R>	</S>	21	50813
	</R></I></R></R>		21	7487
	</R>	</F></F>	18	100458
	</R></P></R>	</S></S></S>	18	6732
	</R></R></I></R>	</S>	18	3391
	</R></S></I></P></R>	</S>	18	8384
</P>	</R></S></I></R>	</S>	18	2311
</P></P></P></P>	</R>	</S></S></S></S>	18	17801
	</R></I></F></R>	</S>	15	928
	</R></I></R>	</S></S></S>	15	66884
	</R></P></R>	</S></S>	15	788
</P>	</R>	</I></S></S></S>	15	2397
	</R></S></S></I></R>	</S></S></S>	15	26888
	</R></S></S></I></R>	</S></S></S></S>	15	2952
	</R></I></P></R>	</S></S></S></S></S>	12	2038
	</R></I></R></I></R>	</S></S></S>	12	1910
	</R></I></R></R>	</S></S>	12	811
	</R></R>	</S></F>	12	19327
	</R></S></I></P></R>	</S></S>	12	113160
	</R></S></S></I></R>	</S>	12	12023
	</R></I></F></R>		12	31595
	</R></I></R></R>	</S>	12	720
	</R></I></R>	</S></S>	12	75599
	</R></R></R>	</S></S></S></S>	12	3815
</P>	</R>	</S></F>	12	4482
</P>	</R></I></P></R>	</S></S>	12	339
</P>	</R></S></I></P></R>	</S></S></S>	12	56468
</P>	</R></S></I></R>	</S></S>	12	5805
</P>	</R>	</F>	12	22198
</P></P>	</R></S></I></R>		12	1419
	</R></R></I></R>		9	1175
	</R></S></R>		9	3937565
	</R></I></P></R>		9	4784
	</R></I></R>	</S></S></S></S></S></S>	9	19
	</R></P></R>	</S></S></S></S>	9	38028
	</R></R>	</S></S></S>	9	22895
</P>	</R></I></P></R>	</S>	9	12419
</P>	</R></R>	</S></S>	9	10205
</P></P></P>	</R>	</S></S></S></S></S></S>	9	6283
</P>	</R></P></R>	</S></S>	9	14200
</P>	</R></S></I></R>	</S></S></S></S>	9	596
</P>	</R></S></S></I></R>	</S></S></S>	9	5725
	</R>	</S></S></F>	6	6535
	</R></F></R>	</S>	6	1075
	</R></I></F></R>	</S></F>	6	581
	</R></I></R></R>	</S></S></S>	6	173
	</R></S></I></R>	</S></S></S></S></S>	6	567

	</R></S></S></I></R>	</S></S>	6	1208
	</R></I></R></I></R>	</S></S></S></S>	6	1223
	</R></I></R>	</S>	6	36975
	</R></P></R>	</S></S></S></S></S></S>	6	23635
	</R></R>	</S></S></S></S></S></S>	6	3148
	</R></S></F></R>	</S>	6	2755
	</R></S></I></R></R>	</S></S></S>	6	303
	</R></S></R>	</S></S>	6	388
	</R></S></S></I></P></R>	</S></S></S>	6	4813
</P>	</R></I></P></R>	</S></S></S></S></S>	6	171
</P>	</R></R>	</S>	6	207
</P></P></P>	</R>	</S></S></S></S></S></S></S>	6	2084
</P></P></P></P>	</R>	</S></S></S>	6	33436
	</R></S></S></I></R>	</S></S></S></S></S>	6	577
</P>	</R>	</S></S></S></S></S></S></S></S>	6	2076
</P>	</R></I></R>		6	883
</P>	</R></R>	</S></S>	6	11133
</P>	</R></S></I></P></R>	</S></S>	6	174
</P>	</R></S></I></R>	</S></S></S></S></S>	6	977
	</R></F></R></I></R>		3	1
	</R></I></P></R>	</S></S></S></S></S></S>	3	138
	</R></I></R>	</S></S></S></S></S></S></S>	3	62
	</R></I></R></I></P></R>	</S></S>	3	243
	</R></I></R></I></R></I></R>	</S></S>	3	17568
	</R></I></R>	</S></S></S></S>	3	2148
	</R></I></S></R>	</S>	3	5
	</R></P></R>		3	326
	</R></R>	</F>	3	1
	</R></R></I></R>	</S></S></S>	3	5287
	</R></R></P></R>	</S>	3	750
	</R></R></R>	</S></S></S>	3	231
	</R></S></I></P></P></R>	</S></S></S>	3	5
	</R></S></I></R>		3	39097
	</R></S></I></R>	</S></S></S></S></S></S></S>	3	223
	</R></S></P></R>	</S></S>	3	317
	</R></S></R>	</F>	3	66293
	</R></S></S></I></P></R>	</S>	3	4244
	</P>	</S></S></S></S></S></S>	3	37
	</R></I></F></R>	</S></S>	3	7197
	</R></I></P></P></R>	</S></S></S>	3	410
	</R></I></P></P></R>	</S></S></S></S>	3	704
	</R></I></R></R>	</S></S></S></S>	3	648
	</R></I></R></S></S></I></P></P></R>	</S></S></S>	3	2
	</R></R></I></R>	</S></S></S></S>	3	191
	</R></R>	</S></S>	3	78

	</R></S></I></P></R>	</S></S></S></S></S>	3	525
	</R></S></I></R></R>	</S>	3	522
	</R></S></R>	</S></S></S></S>	3	4969
	</R></S></R></R>		3	199590
	</R></S></S></I></P></P></R>	</S></S></S>	3	576
	</R></S></S></I></R></R>	</S></S></S></S>	3	89
	</R></S></S></R>	</S></S>	3	5
	</R></S></S></S></I></R>	</S></S></S>	3	4283
	</R></S></S></S></R>	</S></S></S></S>	3	9
	</R></S></S></S></S>		3	4666
	</R></S></S></S></S></I></R>	</S></S></S></S>	3	1408
</P>	</R></S></I></R>		3	1
</P>	</R></S></R>	</S></S>	3	12
</P>	</R></S></S></S></I></R>	</S></S></S>	3	2277
</P></P></P>	</R>		3	30511
	</R></S></S></R>	</S>	3	44
	</R></S></S></S></S></I></P></R>	</S></S></S>	3	11435
</P>	</P></R>	</S></S></S>	3	1037
</P>	</R>	</F></F>	3	1305
</P>	</R></I></R></R>		3	13
</P>	</R></R>	</S></S></S></S></S></S>	3	5
</P>	</R></S></R>	</S>	3	207
</P>	</R></S></S></R>	</S>	3	2199
</P></I>	</R>	</S></S></S></S>	3	81
</P></P>	</R></I></P></R>	</S></S>	3	192
</P></P>	</R></I></R>		3	205
</P></P>	</R></R>	</S>	3	682
</P></P>	</R></S></I></F></R>	</S>	3	54
</P></P>	</R></S></I></R>	</S>	3	3184
</P></P>	</R>	</S></S></S>	3	35
</P></P></P></P>	</R>	</S></S>	3	9149
<b>Σ</b>	<b>197</b>		<b>63634</b>	<b>1124402545</b>

Tab. 24: Vzorce struktury slov podle typů morfémů (preradix. část | radix. č. | postradix. č.) – řazení podle RMSČ

#### 7.2.8.2 Vzorce struktury slov podle typů morfémů (preradix. část | radix. č. | postradix. č.) Řazení podle frekvence v synČNK.

preradix. část	radixová část	postradix. část	frekv. (RMSČ)	frekv. (synČNK)
	</R>		918	239280611
	</R>	</S>	16758	238249152
	</R>	</S></S>	23343	201318117
</P>	</R>	</S></S>	27078	95599497
	</R>	</S></S></S>	23817	82639032
</P>	</R>	</S>	5622	65811445
</P>	</R>	</S></S></S>	18498	53625708

	</R>	</S></S></S></S>	16269	32612036
</P>	</R>	</S></S></S></S>	21021	30315271
</P></P>	</R>	</S></S></S>	2379	9954221
</P></P>	</R>	</S></S>	1947	7960765
	</R>	</S></S></S></S></S>	5067	7082010
</P>	</R>	</S></S></S></S></S>	6357	6644874
</P>	</R>		207	6150939
	</R></R>	</S>	1467	5130491
	</R>	</F>	159	4476068
</P></P>	</R>	</S></S></S></S>	2271	4099518
	</R></S></R>		9	3937565
	</R></I></R>	</S>	1389	3197699
</P></P>	</R>	</S>	579	2622646
	</R></I></R>	</S></S>	2133	2422084
	</R></I></R>	</S></S></S>	1983	2335651
	</R></R>	</S></S>	1188	2153814
</P></P>	</R>	</S></S></S></S></S>	1233	1601094
</P></P>	</R>		48	1416172
	</R></R>	</S></S></S>	756	1376217
</P>	</R></R>		24	1253280
	</R>	</S></F>	129	1170820
	</R></R>	</S></S></S></S>	600	1055782
	</R>	</S></S></S></S></S></S>	939	982036
</P>	</R>	</S></S></S></S></S></S>	1410	836059
</P>	</R></I></R>	</S>	87	593074
	</R></I></R>		51	520331
</P></P></P>	</R>	</S></S></S></S>	156	469202
</P>	</R></I></R>	</S></S></S>	186	444621
	</R></I></R>	</S></S></S></S>	594	383474
	</R></R>		114	382823
</P></P></P>	</R>	</S></S></S>	90	240693
	</R></I></P></R>	</S>	114	226911
	</R></I></P></R>	</S></S></S>	231	222362
</P>	</R></R>	</S></S></S></S>	39	221321
</P></P></P>	</R>	</S></S>	78	207635
</P>	</R></I></R>	</S></S>	108	207113
	</R></S></R></R>		3	199590
	</R></S></I></R>	</S>	135	173473
	</R></S></I></R>	</S></S></S>	342	168572
</P></P>	</R>	</S></S></S></S></S></S>	189	138536
	</R></R></I></R>	</S></S>	45	121871
	</R></S></I></R>	</S></S>	279	115754
	</R></S></I></P></R>	</S></S>	12	113160
	</R></S></I></P></R>	</S></S></S>	51	109054
	</R></I></R></I></R>	</S></S>	36	108243
	</R>	</S></S></S></S></S></S></S>	150	107574
	</R>	</F></F>	18	100458



	</R></R>	</S></S></S></S></S>	111	97575
	</R></I></R>	</S></S>	12	75599
</P>	</R>	</S></S></S></S></S></S>	168	74741
</P>	</R></I></R>	</S></S></S></S>	72	71567
	</R></I></R>	</S></S></S>	15	66884
	</R></S></R>	</F>	3	66293
</P></P></P>	</R>	</S></S></S></S></S>	99	59637
</P>	</R></S></I></P></R>	</S></S></S>	12	56468
	</R></I></P></R>	</S></S>	84	54164
	</R></I></R>	</S></S></S></S></S>	117	53961
	</R></I></R></I></R>	</S>	21	50813
	</R></S></I></R>		3	39097
	</R></P></R>	</S></S></S></S>	9	38028
	</R></I></R>	</S>	6	36975
</P></P></P></P>	</R>	</S></S></S>	6	33436
	</R></I></F></R>		12	31595
</P></P></P>	</R>		3	30511
</P>	</R></R>	</S></S></S></S></S>	33	29140
	</R></S></S></I></R>	</S></S></S>	15	26888
</P>	</R></R>	</S></S></S>	72	25506
	</R></P></R>	</S></S></S></S></S></S>	6	23635
	</R></R>	</S></S></S>	9	22895
</P>	</R>	</F>	12	22198
	</R></R>	</S></F>	12	19327
</P></P></P></P>	</R>	</S></S></S></S>	18	17801
	</R></I></R></I></R></I></R>	</S></S>	3	17568
	</R></S></I></R>	</S></S></S></S>	72	17151
	</R></I></P></R>	</S></S></S></S>	93	16989
</P>	</R></R>	</S>	36	15257
</P>	</R></I></P></R>	</S></S></S></S>	51	14279
</P>	</R></P></R>	</S></S>	9	14200
</P></P></P>	</R>	</S>	27	13990
</P>	</R></S></I></R>	</S></S></S>	45	13377
</P>	</R></I></P></R>	</S>	9	12419
	</R></S></S></I></R>	</S>	12	12023
	</R></S></S></S></S></I></P></R>	</S></S></S>	3	11435
</P>	</R></R>	</S></S>	6	11133
</P>	</R></R>	</S></S>	9	10205
</P></P></P></P>	</R>	</S></S>	3	9149
	</R></S></I></P></R>	</S>	18	8384
</P></P>	</R>	</S></S></S></S></S></S></S>	36	7527
	</R></I></R></R>		21	7487
	</R></R></R>		33	7281
	</R></I></F></R>	</S></S>	3	7197
	</R></P></R>	</S></S></S>	18	6732
</P>	</R></I></R>	</S></S></S></S></S>	27	6590
	</R>	</S></S></F>	6	6535
</P></P></P>	</R>	</S></S></S></S></S></S>	9	6283
	</R></R></R>	</S>	42	5921
</P>	</R></S></I></R>	</S></S>	12	5805

</P>	</R></S></S></I></R>	</S></S></S>	9	5725
	</R></R></I></R>	</S></S></S>	3	5287
	</R></S></R>	</S></S></S></S>	3	4969
	</R></S></R>	</S></S></S>	30	4842
	</R></S></S></I></P></R>	</S></S></S>	6	4813
	</R></I></P></R>		9	4784
	</R></S></S></S></S>		3	4666
</P>	</R>	</S></F>	12	4482
	</R></S></S></S></I></R>	</S></S></S>	3	4283
	</R></S></S></I></P></R>	</S>	3	4244
	</R>	</S></S></S></S></S></S></S></S>	21	4226
	</R></R></R>	</S></S></S></S>	12	3815
	</R></R></I></R>	</S>	18	3391
</P></P>	</R></S></I></R>	</S>	3	3184
	</R></R>	</S></S></S></S></S></S>	6	3148
	</R></R></R>	</S></S>	24	3104
	</R></S></S></I></R>	</S></S></S></S>	15	2952
</P>	</R></I></P></R>	</S></S></S>	27	2834
	</R></S></F></R>	</S>	6	2755
</P>	</R>	</I></S></S></S>	15	2397
</P>	</R></S></I></R>	</S>	18	2311
</P>	</R></S></S></S></I></R>	</S></S></S>	3	2277
</P>	</R></S></S></R>	</S>	3	2199
	</R></I></R>	</S></S></S></S>	3	2148
</P></P></P>	</R>	</S></S></S></S></S></S></S>	6	2084
</P>	</R>	</S></S></S></S></S></S></S></S>	6	2076
	</R></I></P></R>	</S></S></S></S></S>	12	2038
	</R></I></R></I></R>	</S></S></S>	12	1910
	</R></S></R>	</S>	30	1896
</P></P>	</R></S></I></R>		12	1419
	</R></S></S></S></S></I></R>	</S></S></S></S>	3	1408
</P>	</R>	</F></F>	3	1305
	</R></I></R></I></R>	</S></S></S></S>	6	1223
	</R></S></S></I></R>	</S></S>	6	1208
	</R></R></I></R>		9	1175
	</R></F></R>	</S>	6	1075
</P>	</P></R>	</S></S></S>	3	1037
</P>	</R></S></I></R>	</S></S></S></S></S>	6	977
	</R></I></F></R>	</S>	15	928
</P>	</R></I></R>		6	883
	</R></P></R>	</S>	33	854
	</R></I></R></R>	</S></S>	12	811
	</R></P></R>	</S></S>	15	788
	</R></R></P></R>	</S>	3	750
	</R></I></R></R>	</S>	12	720
	</R></I></P></P></R>	</S></S></S></S>	3	704
</P></P>	</R></R>	</S>	3	682

	</R></I></R></R>	</S></S></S></S>	3	648
</P>	</R></S></I></R>	</S></S></S></S>	9	596
	</R></I></F></R>	</S></F>	6	581
	</R></S></S></I></R>	</S></S></S></S></S>	6	577
	</R></S></S></I></P></P></R>	</S></S></S>	3	576
	</R></S></I></R>	</S></S></S></S></S>	6	567
	</R></S></I></P></R>	</S></S></S></S></S>	3	525
	</R></S></I></R></R>	</S>	3	522
	</R></I></P></P></R>	</S></S></S>	3	410
	</R></S></R>	</S></S>	6	388
</P>	</R></I></P></R>	</S></S>	12	339
	</R></P></R>		3	326
	</R></S></P></R>	</S></S>	3	317
	</R></S></I></R></R>	</S></S></S>	6	303
	</R></I></R></I></P></R>	</S></S>	3	243
	</R></R></R>	</S></S></S>	3	231
	</R></S></I></R>	</S></S></S></S></S></S></S>	3	223
</P>	</R></R>	</S>	6	207
</P>	</R></S></R>	</S>	3	207
</P></P>	</R></I></R>		3	205
</P></P>	</R></I></P></R>	</S></S>	3	192
	</R></R></I></R>	</S></S></S></S>	3	191
</P>	</R></S></I></P></R>	</S></S>	6	174
	</R></I></R></R>	</S></S></S>	6	173
</P>	</R></I></P></R>	</S></S></S></S></S>	6	171
	</R></I></P></R>	</S></S></S></S></S></S>	3	138
	</R></S></S></I></R></R>	</S></S></S></S>	3	89
</P></I>	</R>	</S></S></S></S>	3	81
	</R></R>	</S></S>	3	78
	</R></I></R>	</S></S></S></S></S></S></S>	3	62
</P></P>	</R></S></I></F></R>	</S>	3	54
	</R></S></S></R>	</S>	3	44
	</P>	</S></S></S></S></S></S>	3	37
</P></P>	</R>	</S></S></S>	3	35
	</R></I></R>	</S></S></S></S></S></S>	9	19
</P>	</R></I></R></R>		3	13
</P>	</R></S></R>	</S></S>	3	12
	</R></S></S></S></R>	</S></S></S></S>	3	9
	</R></I></S></R>	</S>	3	5
	</R></S></I></P></P></R>	</S></S></S>	3	5
	</R></S></S></R>	</S></S>	3	5
</P>	</R></R>	</S></S></S></S></S></S>	3	5
	</R></I></R></S></S></I></P></P></R>	</S></S></S>	3	2
	</R></F></R></I></R>		3	1
	</R></R>	</F>	3	1

</P>	</R></S></I></R>		3	1
<b>Σ</b>	<b>197</b>		<b>63634</b>	<b>1124402545</b>

Tab. 25: Vzorce struktury slov podle typů morfémů (preradix. část | radix. č. | postradix. č.) – řazení podle RMSČ

### 7.2.8.3 50 nejčastějších vzorců struktury slov podle typů morfémů a slovního druhu (preradix. část | radix. č. | postradix. č.)

Řazení podle frekvence v RMSČ.

pořadí	preradix. část	radixová část	postradix. část	sl. druh	frekv. (RMSČ)	frekv. (synČNK)
1.	</P>	</R>	</S></S>	V	20271	62365765
2.		</R>	</S></S>	S	15870	60431445
3.		</R>	</S>	S	15828	145037783
4.		</R>	</S></S></S>	S	12231	27610395
5.	</P>	</R>	</S></S></S></S>	A	12165	13741359
6.		</R>	</S></S></S></S>	A	8208	21735926
7.	</P>	</R>	</S></S></S>	V	8187	14625329
8.		</R>	</S></S></S>	A	7434	43542252
9.		</R>	</S></S></S></S>	S	6855	9381146
10.	</P>	</R>	</S></S></S></S>	S	5559	13484423
11.		</R>	</S></S>	V	5247	94828439
12.	</P>	</R>	</S></S>	S	5088	19219584
13.	</P>	</R>	</S></S></S>	A	5061	28173343
14.	</P>	</R>	</S>	S	4695	48518319
15.	</P>	</R>	</S></S></S>	S	4311	9387466
16.	</P>	</R>	</S></S></S></S></S>	A	4218	4043437
17.		</R>	</S></S></S></S></S>	A	3351	5636063
18.	</P>	</R>	</S></S></S></S>	V	3096	2680755
19.		</R>	</S></S></S>	V	2781	6029964
20.	</P>	</R>	</S></S></S></S></S>	S	1728	2373387
21.		</R>	</S></S></S></S></S>	S	1491	1375105
22.		</R>	</S></S>	C	1455	7546338
23.	</P>	</R>	</S></S>	C	1440	6226374
24.		</R></R>	</S>	S	1356	2993348
25.		</R></I></R>	</S></S></S>	A	1338	1601545
26.	</P></P>	</R>	</S></S>	V	1305	5377650
27.		</R>	</S></S></S>	C	1287	2238993
28.		</R></I></R>	</S>	S	1272	2140200
29.	</P></P>	</R>	</S></S></S></S>	A	1221	2946211
30.		</R></I></R>	</S></S>	S	1098	1258029
31.	</P>	</R>	</S></S></S></S></S></S>	A	1077	555318
32.		</R></R>	</S></S>	S	1065	1814550
33.	</P>	</R>	</S></S></S>	C	921	1435295
34.	</P>	</R>	</S>	C	879	14075501
35.		</R>	</S></S></S></S>	V	873	996402
36.	</P></P>	</R>	</S></S></S>	A	816	7203156
37.	</P></P>	</R>	</S></S></S></S></S>	A	762	1158865
38.		</R></I></R>	</S></S>	A	744	672229

39.	</P></P>	</R>	</S></S></S>	V	735	1944722
40.		</R>	</S></S>	A	708	20245877
41.		</R>	</S>	C	687	18958063
42.	</P></P>	</R>	</S></S></S></S>	S	687	450936
43.	</P></P>	</R>	</S></S></S>	S	657	536947
44.		</R>	</S></S></S></S></S></S>	A	621	673913
45.		</R></I></R>	</S></S></S>	S	606	330903
46.		</R></R>	</S></S></S></S>	A	555	1011118
47.		</R>		I	429	286550
48.	</P></P>	</R>	</S></S></S></S></S>	S	405	403765
49.	</P></P>	</R>	</S></S>	C	381	1655157
50.		</R></I></R>	</S></S></S></S>	A	378	242574
...	...	...	...	...	...	...
Σ		446			63634	1124402545

Tab. 26: 50 nejčastějších vzorců struktury slov podle typů morfémů a slovního druhu (preradix. část | radix. č. | postradix. č.) – řazení podle RMSČ

#### 7.2.8.4 50 nejčastějších vzorců struktury slov podle typů morfémů a slovního druhu (preradix. část | radix. č. | postradix. č.)

Řazení podle frekvence v synČNK.

pořadí	preradix. část	radixová část	postradix. část	sl. druh	frekv. (RMSČ)	frekv. (synČNK)
1.		</R>	</S>	S	15828	145037783
2.		</R>		R	114	133344458
3.		</R>	</S></S>	V	5247	94828439
4.		</R>		K	102	74508181
5.	</P>	</R>	</S></S>	V	20271	62365765
6.		</R>	</S>	P	78	61054764
7.		</R>	</S></S>	S	15870	60431445
8.	</P>	</R>	</S>	S	4695	48518319
9.		</R>	</S></S></S>	A	7434	43542252
10.	</P>	</R>	</S></S></S>	A	5061	28173343
11.		</R>	</S></S></S>	S	12231	27610395
12.		</R>		C	213	25536448
13.		</R>	</S></S></S></S>	A	8208	21735926
14.		</R>	</S></S>	A	708	20245877
15.	</P>	</R>	</S></S>	S	5088	19219584
16.		</R>	</S>	C	687	18958063
17.		</R>	</S></S>	P	36	17624097
18.	</P>	</R>	</S></S></S>	V	8187	14625329
19.	</P>	</R>	</S>	C	879	14075501
20.	</P>	</R>	</S></S></S></S>	A	12165	13741359
21.	</P>	</R>	</S></S></S></S>	S	5559	13484423
22.		</R>	</S>	N	57	12405362
23.	</P>	</R>	</S></S></S>	S	4311	9387466
24.		</R>	</S></S></S></S>	S	6855	9381146
25.		</R>	</S></S>	C	1455	7546338
26.	</P></P>	</R>	</S></S></S>	A	816	7203156
27.	</P>	</R>	</S></S>	A	270	6804358

28.	</P>	</R>	</S></S>	C	1440	6226374
29.		</R>	</S></S></S>	V	2781	6029964
30.		</R>	</S></S></S></S></S>	A	3351	5636063
31.	</P></P>	</R>	</S></S>	V	1305	5377650
32.	</P>	</R>	</S></S></S></S></S>	A	4218	4043437
33.		</R></S></R>		P	3	3891684
34.	</P>	</R>		C	195	3530522
35.		</R>		P	9	3398289
36.		</R>	</F>	P	21	3053442
37.		</R></R>	</S>	S	1356	2993348
38.	</P></P>	</R>	</S></S></S></S>	A	1221	2946211
39.	</P>	</R>	</S></S></S></S>	V	3096	2680755
40.	</P>	</R>		R	9	2618809
41.		</R>	</S></S></S>	P	12	2616182
42.	</P>	</R>	</S></S></S></S></S>	S	1728	2373387
43.		</R>	</S></S></S>	C	1287	2238993
44.		</R></I></R>	</S>	S	1272	2140200
45.		</R>		T	48	2037431
46.	</P></P>	</R>	</S></S></S>	V	735	1944722
47.	</P></P>	</R>	</S>	S	351	1822722
48.		</R></R>	</S></S>	S	1065	1814550
49.	</P></P>	</R>	</S></S>	C	381	1655157
50.		</R></I></R>	</S></S></S>	A	1338	1601545
...	...	...	...	...	...	...
Σ		446			63634	1124402545

Tab. 27: 50 nejčastějších vzorců struktury slov podle typů morfémů a slovního druhu (preradix. část | radix. č. | postradix. č.) – řazení podle synČNK

#### 7.2.8.5 50 nejčastějších vzorců struktury slov podle celé značky s vynecháním slovního druhu (preradix. část | radix. č. | postradix. č.)

Řazení podle frekvence v RMSČ.

pořadí	preradix. část	radixová část	postradix. část	frekv. (RMSČ)	frekv. (synČNK)
1.	7	<R ... 0></R>	<S ... k></S><S ... g></S>	20208	61678631
2.		<R ... 0></R>	<S ... g></S>	15825	145037782
3.		<R ... 0></R>	<S ... s></S><S ... g></S>	14673	49213063
4.	<P ...></P>	<R ... 0></R>	<S ... k></S><S ... g></S><S ... k></S><S ... g></S>	9777	6836426
5.		<R ... 0></R>	<S ... s></S><S ... s></S><S ... g></S>	9441	18147143
6.		<R ... 0></R>	<S ... s></S><S ... k></S><S ... g></S>	7245	42611493
7.	<P ...></P>	<R ... 0></R>	<S ... k></S><S ... k></S><S ... g></S>	5976	9956750
8.		<R ... 0></R>	<S ... s></S><S ... s></S><S ... k></S><S ... g></S>	5655	16459202
9.		<R ... 0></R>	<S ... k></S><S ... g></S>	5241	94780666
10.	<P ...></P>	<R ... 0></R>	<S ... g></S>	4401	43821573
11.	<P ...></P>	<R ... 0></R>	<S ... s></S><S ... g></S>	3762	12705690
12.	<P ...></P>	<R ... 0></R>	<S ... s></S><S ... k></S><S ... g></S>	3507	25080021
13.	<P ...></P>	<R ... 0></R>	<S ... k></S><S ... g></S><S ... k></S><S ... g></S>	2667	8593581
14.	<P ...></P>	<R ... 0></R>	<S ... s></S><S ... s></S><S ... g></S>	2550	5066267
15.		<R ... 0></R>	<S ... k></S><S ... k></S><S ... g></S>	2031	4761821
16.	<P ...></P>	<R ... 0></R>	<S ... s></S><S ... s></S><S ... k></S><S ... g></S>	1980	1677789
17.		<R ... 0></R>	<S ... s></S><S ... s></S><S ... k></S><S ... g></S>	1869	3148224
18.		<R ... 0></R>	<S ... k></S><S ... g></S><S ... k></S><S ... g></S>	1848	4002489

19.		<R ... 0></R>	<S ... s></S><S ... s></S><S ... s></S><S ... g></S>	1848	1767800
20.		<R ... 0></R>	<S ... k:s></S><S ... g:s></S><S ... k></S><S ... g></S>	1812	2667959
21.	<P ...></P>	<R ... 0></R>	<S ... k:s></S><S ... g:s></S><S ... s></S><S ... g></S>	1434	1973603
22.	<P ...></P>	<R ... 0></R>	<S ... k:s></S><S ... k:s></S><S ... g:s></S><S ... k></S><S ... g></S>	1419	836373
23.	<P ...></P>	<R ... 0></R>	<S ... s></S><S ... s></S><S ... k></S><S ... g></S>	1398	4817590
24.		<R ... 0></R>	<S ... s></S><S ... g:s></S>	1392	7375625
25.		<R ... 12></R><R ... 22></R>	<S ... g></S>	1356	2993348
26.	<P ...></P>	<R ... 0></R>	<S ... g:s></S><S ... k></S><S ... g></S>	1323	1630291
27.	<P ...></P>	<R ... 0></R>	<S ... s></S><S ... g:s></S>	1254	5679018
28.	<P ...></P>	<R ... 0></R>	<S ... s></S><S ... k></S><S ... g></S>	1239	2279631
29.	<P ...></P><P ...></P>	<R ... 0></R>	<S ... k></S><S ... g></S>	1209	5057365
30.		<R ... 0></R>	<S ... k:s></S><S ... s></S><S ... g></S>	1173	5227415
31.		<R ... 0></R>	<S ... k:s></S><S ... g:s></S><S ... s></S><S ... g></S>	1161	1298652
32.	<P ...></P><P ...></P>	<R ... 0></R>	<S ... k:s></S><S ... g:s></S><S ... k></S><S ... g></S>	990	2114541
33.		<R ... 12></R><R ... 22></R>	<S ... s></S><S ... g></S>	978	1707822
34.		<R ... 0></R>	<S ... s></S><S ... s></S><S ... k></S><S ... g></S>	969	1026417
35.		<R ... 12></R><I ... s></I><R ... 22></R>	<S ... g></S>	921	1455702
36.		<R ... 0></R>	<S ... s></S><S ... k></S><S ... g></S>	888	2320986
37.		<R ... 0></R>	<S ... s></S><S ... s></S><S ... g:s></S>	870	1290387
38.		<R ... 0></R>	<S ... k></S><S ... g></S>	846	6095943
39.		<R ... 12></R><I ... s></I><R ... 22></R>	<S ... s></S><S ... k></S><S ... g></S>	840	1119061
40.	<P ...></P>	<R ... 0></R>	<S ... s></S><S ... k></S><S ... k></S><S ... g></S>	810	665884
41.	<P ...></P>	<R ... 0></R>	<S ... g:s></S>	801	12491946
42.		<R ... 0></R>	<S ... k></S><S ... g></S>	708	20245877
43.		<R ... 12></R><I ... s></I><R ... 22></R>	<S ... s></S><S ... g></S>	693	415463
44.	<P ...></P><P ...></P>	<R ... 0></R>	<S ... s></S><S ... k></S><S ... g></S>	675	6103835
45.	<P ...></P>	<R ... 0></R>	<S ... s></S><S ... k:s></S><S ... g:s></S><S ... k></S><S ... g></S>	666	359282
46.		<R ... 0></R>	<S ... g:s></S>	657	16662712
47.		<R ... 0></R>	<S ... s></S><S ... k></S><S ... g></S>	657	750383
48.	<P ...></P>	<R ... 0></R>	<S ... k:s></S><S ... s></S><S ... s></S><S ... k></S><S ... g></S>	651	1043562
49.	<P ...></P>	<R ... 0></R>	<S ... k:s></S><S ... s></S><S ... g></S>	618	976846
50.		<R ... 0></R>	<S ... s></S><S ... k></S><S ... k></S><S ... g></S>	606	382253
...	...	...	...	...	...
Σ		1380		63634	1124402545

Tab. 28: 50 nejčastějších vzorců struktury slov podle celé značky s vynecháním slovního druhu (preradix. část | radix. č. | postradix. č.) – řazení podle RMSČ

#### 7.2.8.6 50 nejčastějších vzorců struktury slov podle celé značky s vynecháním slovního druhu (preradix. část | radix. č. | postradix. č.)

Řazení podle frekvence v synČNK.

pořadí	preradix. část	radixová část	postradix. část	frekv. (RMSČ)	frekv. (synČNK)
1.		<R ... 0></R>	<S ... g></S>	15825	145037782
2.		<R ... 0></R>		114	133344458
3.		<R ... 0></R>	<S ... k></S><S ... g></S>	5241	94780666

4.		<R ... 0></R>		102	74508181
5.	<P ...></P>	<R ... 0></R>	<S ... k></S><S ... g></S>	20208	61678631
6.		<R ... 0></R>	<S ... g></S>	78	61054764
7.		<R ... 0></R>	<S ... s></S><S ... g></S>	14673	49213063
8.	<P ...></P>	<R ... 0></R>	<S ... g></S>	4401	43821573
9.		<R ... 0></R>	<S ... s></S><S ... k></S><S ... g></S>	7245	42611493
10.		<R ... 0></R>		213	25536448
11.	<P ...></P>	<R ... 0></R>	<S ... s></S><S ... k></S><S ... g></S>	3507	25080021
12.		<R ... 0></R>	<S ... k></S><S ... g></S>	708	20245877
13.		<R ... 0></R>	<S ... s></S><S ... s></S><S ... g></S>	9441	18147143
14.		<R ... 0></R>	<S ... g:s></S>	657	16662712
15.		<R ... 0></R>	<S ... s></S><S ... s></S><S ... k></S><S ... g></S>	5655	16459202
16.	<P ...></P>	<R ... 0></R>	<S ... s></S><S ... g></S>	3762	12705690
17.	<P ...></P>	<R ... 0></R>	<S ... g:s></S>	801	12491946
18.		<R ... 0></R>	<S ... k></S><S ... g></S>	12	10878715
19.		<R ... 0></R>	<S ... g></S>	39	10758219
20.	<P ...></P>	<R ... 0></R>	<S ... k></S><S ... k></S><S ... g></S>	5976	9956750
21.	<P ...></P>	<R ... 0></R>	<S ... k:s></S><S ... g:s></S><S ... k></S><S ... g></S>	2667	8593581
22.		<R ... 0></R>	<S ... s></S><S ... g:s></S>	1392	7375625
23.	<P ...></P>	<R ... 0></R>	<S ... k:s></S><S ... g:s></S><S ... k></S><S ... g></S>	9777	6836426
24.	<P ...></P>	<R ... 0></R>	<S ... k></S><S ... g></S>	270	6804358
25.	<P ...></P><P ...></P>	<R ... 0></R>	<S ... s></S><S ... k></S><S ... g></S>	675	6103835
26.		<R ... 0></R>	<S ... k></S><S ... g></S>	846	6095943
27.	<P ...></P>	<R ... 0></R>	<S ... s></S><S ... g:s></S>	1254	5679018
28.		<R ... 0></R>	<S ... k:s></S><S ... s></S><S ... g></S>	1173	5227415
29.	<P ...></P>	<R ... 0></R>	<S ... s></S><S ... s></S><S ... g></S>	2550	5066267
30.	<P ...></P><P ...></P>	<R ... 0></R>	<S ... k></S><S ... g></S>	1209	5057365
31.	<P ...></P>	<R ... 0></R>	<S ... s></S><S ... s></S><S ... k></S><S ... g></S>	1398	4817590
32.		<R ... 0></R>	<S ... k></S><S ... k></S><S ... g></S>	2031	4761821
33.		<R ... 0></R>	<S ... s v></S><S ... g></S>	255	4712275
34.		<R ... 0></R>	<S ... g></S><S ... g></S>	6	4622482
35.	<P ... v></P>	<R ... 0></R>	<S ... g></S>	288	4308163
36.		<R ... 0></R>	<S ... k:s></S><S ... g:s></S><S ... k></S><S ... g></S>	1848	4002489
37.		<R ... 12></R><R ... s></S><R ... 22></R>		3	3891684
38.		<R ... 0></R>		9	3398289
39.		<R ... 0></R>	<S ... s></S><S ... s></S><S ... s></S><S ... k></S><S ... g></S>	1869	3148224
40.		<R ... 0></R>	<F ... g></F>	21	3053442
41.		<R ... 12></R><R ... 22></R>	<S ... g></S>	1356	2993348
42.	<P ...></P>	<R ... 0></R>		186	2965152
43.		<R ... 0></R>	<S ... k:s></S><S ... g:s></S><S ... k></S><S ... g></S>	1812	2667959
44.	<P ...></P>	<R ... 0></R>	<S ... k></S><S ... g></S>	570	2652458
45.	<P ...></P>	<R ... 0></R>		9	2618809
46.		<R ... 0></R>	<S ... g:s></S><S ... g:s></S><S ... g></S>	3	2588928
47.		<R ... 0></R>	<S ... s></S><S ... k></S><S ... g></S>	888	2320986



48.	<P ...></P>	<R ... 0></R>	<S ... s></S><S ... k></S><S ... g></S>	1239	2279631
49.	<P ...></P><P ...></P>	<R ... 0></R>	<S ... k:s></S><S ... g:s></S><S ... k></S><S ... g></S>	990	2114541
50.		<R ... 0></R>		48	2037431
...	...	...	...	...	...
Σ		1380		63634	1124402545

Tab. 29: 50 nejčastějších vzorců struktury slov podle celé značky s vynecháním slovního druhu (preradix. část | radix. č. | postradix. č.) – řazení podle synČNK

### 7.2.8.7 50 nejčastějších vzorců struktury slov podle formy, celé značky a slovního druhu (preradix. část | postradix. č.)

Řazení podle frekvence v RMSČ.

pořadí	preradix. část	postradix. část	sl. druh	frekv. (RMSČ)	frekv. (synČNK)
1.		<S S g>Ø</S>	S	7996	87548248
2.		<S S g>a</S>	S	3020	38642010
3.		<S A s>ov</S><S A k>Ø</S><S A g>ý</S>	A	2026	5451443
4.		<S S s>k</S><S S g>a</S>	S	1622	5709136
5.		<S A s>n</S><S A k>Ø</S><S A g>ý</S>	A	1452	8054716
6.		<S V k>ov</S><S V k>a</S><S V g>ti</S>	V	1440	4949008
7.		<S V k>i</S><S V g>ti</S>	V	1182	11634496
8.		<S V k>a</S><S V g>ti</S>	V	1074	7038712
9.		<S A s>n</S><S A k>Ø</S><S A g>i</S>	A	1016	12010473
10.		<S A s>sk</S><S A k>Ø</S><S A g>ý</S>	A	1006	7170971
11.		<S A k>Ø</S><S A g>ý</S>	A	916	18594441
12.		<S S s>ek</S><S S g>Ø</S>	S	884	2914233
13.		<S S s>i</S><S S g>e</S>	S	760	4552830
14.		<S A s>i</S><S A s>k</S><S A k>Ø</S><S A g>ý</S>	A	732	4032221
15.		<S C s>n</S><S C g>s>ě</S>	C	662	5337288
16.		<S S g>e</S>	S	626	9254559
17.		<S V k>nou</S><S V g>ti</S>	V	614	1057137
18.		<S S s>n</S><S S s>ost</S><S S g>Ø</S>	S	566	1697790
19.		<S S g>o</S>	S	560	12319333
20.		<S S k>s>e</S><S S g>s>n</S><S S k>i</S><S S g>Ø</S>	S	502	2178220
21.		<S S s>ák</S><S S g>Ø</S>	S	488	1273212
22.	<P V>vy</P>	<S V k>a</S><S V g>ti</S>	V	488	2485566
23.	<P V>za</P>	<S V k>a</S><S V g>ti</S>	V	484	984860
24.	<P V>vy</P>	<S V k>i</S><S V g>ti</S>	V	462	2246442
25.	<P V>vy</P>	<S V k>ov</S><S V k>a</S><S V g>ti</S>	V	446	1082034
26.		<S A k>s>ov</S><S A k>s>a</S><S A g>s>n</S><S A k>Ø</S><S A g>ý</S>	A	432	663084
27.	<P V>za</P>	<S V k>i</S><S V g>ti</S>	V	420	2201873
28.	<P V>za</P>	<S V k>ov</S><S V k>a</S><S V g>ti</S>	V	416	626035
29.		<S S k>s>á</S><S S g>s>n</S><S S k>i</S><S S g>Ø</S>	S	414	975064
30.		<S S s>ec</S><S S g>Ø</S>	S	408	2012878
31.		<S A k>s>a</S><S A g>s>v</S><S A k>Ø</S><S A g>ý</S>	A	392	116596
32.		<S S s>ič</S><S S s>k</S><S S g>a</S>	S	390	555363
33.		<S S s>ac</S><S S g>e</S>	S	376	2792125
34.		<S S s>ik</S><S S g>Ø</S>	S	372	949988
35.		<S A k>s>i</S><S A g>s>v</S><S A k>Ø</S><S A g>ý</S>	A	368	226251

36.		<S S s>in</S><S S g>a</S>	S	368	2262336
37.	<P V>na</P>	<S V k>a</S><S V g>ti</S>	V	364	657325
38.		<S A s>ov</S><S A s>it</S><S A k>Ø</S><S A g>ý</S>	A	360	50146
39.		<S S s>n</S><S S s>ík</S><S S g>Ø</S>	S	360	1647755
40.		<S S s>ist</S><S S g>a</S>	S	348	1650715
41.	<P V>z</P>	<S V k>i</S><S V g>ti</S>	V	344	2328305
42.	<P A>vy</P>	<S A k>s>e</S><S A g>s>n</S><S A k>Ø</S><S A g>ý</S>	A	332	403039
43.	<P V>po</P>	<S V k>a</S><S V g>ti</S>	V	320	2380615
44.	<P V>roz</P>	<S V k>a</S><S V g>ti</S>	V	314	125802
45.	<P V>po</P>	<S V k>i</S><S V g>ti</S>	V	312	2656219
46.	<P V>na</P>	<S V k>i</S><S V g>ti</S>	V	304	1281444
47.		<S S s>ic</S><S S g>e</S>	S	302	1336457
48.	<P V>u</P>	<S V k>i</S><S V g>ti</S>	V	296	1241026
49.	<P A>za</P>	<S A k>s>e</S><S A g>s>n</S><S A k>Ø</S><S A g>ý</S>	A	288	409852
50.	<P V>o</P>	<S V k>i</S><S V g>ti</S>	V	270	1302135
...	...	...	...	...	...
Σ		12038		63634	1124402545

Tab. 30: 50 nejčastějších vzorců struktury slov podle formy, celé značky a slovního druhu (preradix. část / postradix. č.) – řazení podle RMSČ

#### 7.2.8.8 50 50 nejčastějších vzorců struktury slov podle formy, celé značky a slovního druhu (preradix. část / postradix. č.)

Řazení podle frekvence v synČNK.

pořadí	preradix. část	postradix. část	sl. druh	frekv. (RMSČ)	frekv. (synČNK)
1.			R	76	133344458
2.		<S S g>Ø</S>	S	7996	87548248
3.			K	76	74688476
4.		<S V k>Ø</S><S V g>ti</S>	V	114	47118479
5.		<S P g>Ø</S>	P	20	41479460
6.		<S S g>a</S>	S	3020	38642010
7.			C	222	26101825
8.		<S A k>Ø</S><S A g>ý</S>	A	916	18594441
9.		<S P g>en</S>	P	12	14409475
10.		<S S g>o</S>	S	560	12319333
11.		<S A s>n</S><S A k>Ø</S><S A g>í</S>	A	1016	12010473
12.		<S V k>i</S><S V g>ti</S>	V	1182	11634496
13.		<S V k>í</S><S V g>ti</S>	V	74	11099602
14.		<S P k>Ø</S><S P g>ý</S>	P	30	10908946
15.		<S S g>e</S>	S	626	9254559
16.		<S A s>n</S><S A k>Ø</S><S A g>ý</S>	A	1452	8054716
17.		<S A s>k</S><S A k>Ø</S><S A g>ý</S>	A	194	8053967
18.		<S V k>Ø</S><S V g>i</S>	V	44	7425365
19.			P	14	7413380
20.		<S A s>sk</S><S A k>Ø</S><S A g>ý</S>	A	1006	7170971
21.		<S V k>a</S><S V g>ti</S>	V	1074	7038712
22.		<S N g>Ø</S>	N	52	6688989
23.	<P A>ne</P>	<S A k>Ø</S><S A g>ý</S>	A	50	5974835
24.		<S S s>k</S><S S g>a</S>	S	1622	5709136

25.		<S A s>ov</S><S A k>Ø</S><S A g>ý</S>	A	2026	5451443
26.	<P A>ne</P>	<S A s>k</S><S A k>Ø</S><S A g>ý</S>	A	8	5411664
27.		<S C s>n</S><S C g>s>ě</S>	C	662	5337288
28.		<S V k>ov</S><S V k>a</S><S V g>ti</S>	V	1440	4949008
29.		<S C g>s>e</S>	C	164	4779359
30.		<S P g>ůj</S><S P g>Ø</S>	P	4	4622482
31.		<S S s>i</S><S S g>e</S>	S	760	4552830
32.	<P S>zá</P>	<S S g>Ø</S>	S	160	4395026
33.		<S A s>ic</S><S A s>k</S><S A k>Ø</S><S A g>ý</S>	A	732	4032221
34.		<S V k>e</S><S V g>ti</S>	V	178	3948742
35.	<P S>po</P>	<S S g>Ø</S>	S	192	3668726
36.		<S N g>a</S>	N	4	3089073
37.		<S P g>o</S>	P	42	3084707
38.	<P A>ne</P>	<S A s>n</S><S A k>Ø</S><S A g>ý</S>	A	162	3068324
39.		<F P g>ž</F>	P	4	2932030
40.		<S S s>ek</S><S S g>Ø</S>	S	884	2914233
41.		<S S s>b</S><S S g>a</S>	S	142	2833788
42.		<S S s>ac</S><S S g>e</S>	S	376	2792125
43.		<S V k>á</S><S V g>ti</S>	V	56	2756587
44.	<P S>ná</P>	<S S g>Ø</S>	S	140	2694169
45.	<P V>po</P>	<S V k>i</S><S V g>ti</S>	V	312	2656219
46.		<S V k>ě</S><S V g>ti</S>	V	136	2641619
47.		<S C g>s>ě</S>	C	152	2600400
48.		<S P g>s>ech</S><S P g>s>en</S><S P g>Ø</S>	P	2	2588928
49.	<P R>po</P>		R	2	2527262
50.	<P V>vy</P>	<S V k>a</S><S V g>ti</S>	V	488	2485566
...	...	...	...	...	...
Σ		12038		63634	1124402545

Tab. 31: 50 nejčastějších vzorců struktury slov podle formy, celé značky a slovního druhu (preradix. část / postradix. č.) – řazení podle synČNK

## 7.2.9 Morfematické alternace

Z naznačených důvodů se zde nezabývám problematikou alomorfie, mohu však alespoň naznačit, jaké části slovní zásoby se týká. Na základě podrobného rozboru flektivních paradigmát jsem dospěl ke zjištění, že na celkový počet 63 676 hesel připadá 32 941 případů grafematické alternace bez započítání opakovaných hodnot. U 14 175 případů variuje předposlední morfém ve slově, u 8 684 hesel třetí od konce, čtvrtý potom u 3 221 základních jednotek. Prefix má nějaké alternace v 6 861 případech, s výjimkou 57 slov (u nichž jde nejčastěji o vokalizovanou a nevokalizovanou alternantu při konjugaci) se všechny týkají derivování superlativních tvarů adjektiv a adverbíí.

## 7.3 Predikce kvantitativních charakteristik propriálního systému češtiny

Podle tradičních systematik (např. ŠRÁMEK 1999, ESČ aj.) se vlastní jména dělí takto:

**I. BIONYMA** (vlastní jm. živých (jakoby živých) bytostí a organismů fungujících ve společenských vztazích)

- 1) **antroponyma** (jm. osobní)
- 2) **nepravá antroponyma** (smyšlená apod., např. *Meluzína*)
- 3) **zoonyma** (skutečných živočichů)
- 4) **nepravá zoonyma** (*kůň Pegas*)

### 5) **fytonyma** (rostlin)

## II. **ABIONYMA** (neživých objektů a jevů)

### 1) **toponyma** (vl. jm. zeměpisná / geografická = objekty v krajině, kartograficky fixované)

A) **anoikonyma** (pomístní jména / mikrotoponymum / jm. vl. nesídlíšní = název neobydleného objektu, jm. neživého, pevně fixovaného přírodního objektu a jevu na Zemi)

a) **hydronyma** (vod: jm. moří, zálivů, průlivů, jezer, bažin, rybníků, přehrad, studánek, pramenů, řek, potoků, peřejí, vodopádů, průplavů aj.)

b) **oronyma** (vl. jm. tvarů vertikální členitosti zemského povrchu i mořského dna: pohoří, vrchovin, jednotlivých hor, vrcholků, hřbetů, kopců, skal, údolí, dolin, průsmyků, závrťů, sedel, propastí, nížin, rovin, tabulí, pánví aj.)

c) **agronyma** (vl. jm. pozemková / traťová = jednotlivých pozemků, jejich seskupení, a to obdělávaných: zemědělský hon, pole, louka, pastvina, zahrada, vinice atd. i neobdělávaných n. lesních: polesí, les, obora aj.)

d) **hodonyma** (vl. jm. dopravních cest: jména silnic, dálnic, cest, tunelů, brodů, mostů, přívozů, železnic, lanovek atd.)

e) **urbanonyma** (vl. jm. ulic, náměstí a jiných veřejných prostranství)

f) vl. jm. balvanů, kapliček, božích muk, křížků, hřbitovů, pomníků, mohyl, lomů, dolů, rozhleden, trigonometrických bodů, ukazatelů cest atp.

B) **oikonyma** (místní jména, jm. vl. sídelní)

2) **kosmonyma** / **astronyma** (vl. jm. astronomická = objekty mimo Zemi, tj. galaxií, planet, hvězd, těles aj.)

3) **chrématonyma** (lidských výtvorů nezakotvených v přírodě: tavený sýr *Apetito*, kniha *Babička*)

A) **jm. společenských jevů**

a) **chrononyma** / časové pojmy (svátky, epochy)

b) **akcionyma** / dějinné události (*Pražské povstání*)

c) **faleronyma** / společenská ocenění, vyznamenání, tituly (*Řád bílého lva*)

B) **ergonyma, institucionyma** / **jm. společenských institucí** (*Kruh přátel českého jazyka*)

a) vojenských, politických (*Občanská demokratická strana*)

b) správních, zákonodárných, administrativních (*Parlament ČR*)

c) hospodářských, průmyslových, obchodních (*Svit, Česká národní banka*)

d) školských, kulturních, vědeckých, zdravotnických (*Národní divadlo, Univerzita Karlova*)

C) **jména výsledků společenské činnosti**

a) **aktonyma, dokumentonyma** / diplomatických aktů a dokumentů (*Listina základních lidských práv a svobod*)

b) **ideonyma** / umělecké činnosti (*Má vlast*)

c) jedinečných výrobků či předmětů (*Titanik*)

d) **porejonyma** / dopravních spojů (*rychlík Krakonoš*)

e) **pragmatonyma** / průmyslových výrobků a polnohospodářsky vypěstovaných plemen, odrůd (*kolo Favorit, hodinky Prim, jablko Sparťan*)

Po uskutečnění jen několika sond, které pomohly naznačit kvantitativní rozměry onymických systémů, docházím k vcelku očekávanému zjištění, že poměr apelativ a proprií je přinejmenším srovnatelný, při započítání i víceslovných jednotek dokonce vlastní jména zřejmě převládají. Tak např. křestních jmen eviduje registr obyvatel ČR pro muže 61 923 různých, včetně víceslovných, 26 972 jedinečných a 42 149 kombinací. Celkový nositelů je 5,27 milionu, nejčastějších 10 jmen představuje formálně-morfologicky velmi pestrá skupinu

(Jiří, Jan, Petr, Josef, Pavel, Jaroslav, Martin, Tomáš, Miroslav, František), o jejich morfematickém členění lze kromě deklinačního morfu uvažovat snad jen u jmen Jar-o-slav, Mir-o-slav a Frant-iš-ek. V RMSČ je jediný František, a to bez segmentace (František-Ø). Bez spolehlivých etymologických údajů je jakékoliv členění (mimo gramatický morfém) až na drobné výjimky plné nástrah (v podobě lidových etymologií apod.). Delimitace vedené formálněmorfologickou snahou o začlenění do (počítačového) morfologického systému – nejlépe založeného na morfematických jednotkách (viz kap. 8) – však smysluplná a relativně bezpečná je. V intencích druhé kapitoly na počátku vznikne velké množství neuronových okruhů, zde málo produktivních (mnohdy monofunkčních) kořenů, které si však postupným prohlubováním (vytvářením) znalostí, resp. pozvolným nárůstem motivovanosti začne vytvářet hustší a hustší síť synapsí, takže se nakonec objem periferních radixů (příp. jiných morfémů) zřejmě sníží.

Na prvotní expanzi se může dále např. podílet asi 186 tisíc mužských, resp. cca 209 tisíc ženských příjmení (převážně) rodilých mluvčí češtiny, stejně jako 4949 různých částí názvů obcí a měst, příp. 22 402 komponent víceslovných názvů ulic, 63 465 složek místních nebo na 50 tisíc pomístních jmen (z Čech). To je jen velmi hrubý náčrt kvantitativních souvislostí, jen velmi malá část propriálního systému rekapitulovaného výše. Čekat však, až se s tímto vpravdě gigantickým tématem po svém vypořádají specialisté, onomastici, je navýsost pošetilé, protože mimo dohled dnes žijících generací českých mluvčích. Počítačový morfematický analyzátor (syntetizér) češtiny by měl vzniknout podstatně dříve. Jak by měl přibližně vypadat, ukáže následující kapitola.

#### **7.4 Shrnutí**

Obsahem sedmé kapitoly je pokus o přiblížení základních kvantitativních charakteristik apelativní části morfematického systému češtiny. Prototypem systému se stal mnohostranně zhodnocený obsah Retrográdního morfematického slovníku češtiny E. Slavíčkové. Z něj vyvozené statistické přehledy nejčastějších typů morfémů a jejich kombinatoriky – nahlížené z hlediska slabičných, fonematických, slovnědruhových a zčásti též stylistických aspektů – byly konfrontovány s kvantitativními údaji z největších českých (synchronních) korpusů psané češtiny, a to v úhrnném objemu 1,2 miliardy textových slov. Při porovnání hesel slovníku s lemmaty z korpusů se zdá být až 93 % textů spolehlivě morfematicky členitelných pouze s využitím dosavadních poznatků.

## 8 Návrh efektivní morfemické syntézy

### 8.0 Motto

Ockhamova „břitva“: *Nomina non sunt multiplicanda, praeter necessitatem*<sup>88</sup>

### 8.1 Obecná východiska

V předcházejících kapitolách jsem porůznu dospíval k rozmanitým podnětům, jejichž využití bylo schraňováno pro neurčitě pojmenovávanou počítačovou reprezentaci morfemické struktury češtiny, případně morfemickou analýzu a syntézu. V druhé kapitole jsem mezi inspirativní návrhy vycházející z přehledu vybraných poznatků kognitivních věd zahrnul nerozlišovanou pozornost věnovanou psané i mluvené (transkribované) podobě jazyka, modulární rozdělení na znalosti a pravidla, a výhody formální aktivovanosti na úrovni morfémů. Již v první kapitole jsem zase opakovaně vyjadřoval pochybnosti o možnosti exaktního stanovení formy, která je přesně nositelem (vehikulem) gramémů, příp. sémémů, a uvažoval spíše o sémantických distinktivních rysech. Téma páté kapitoly mě přimělo k rezultativnímu prohlášení, že se morfemická analýza musí po vzoru morfologické analýzy plně algoritmizovat. V šesté kapitole jsem si začal sám odporovat, když jsem konstatoval, že žádné taxativní vymezení zásad pro automatickou delimitaci morfémů nepokládám za možné, a tedy i smysluplné. Naopak jsem nastínil představu dynamického systému, který by mimo jiné měl dokázat zabránit nekonzistencím a závislosti na jedné konkrétní znalostní bázi či teorii. Rád bych zde dokázal, že uvedené kontradikce jsou pouze zdánlivé. V samotném výkladu i v popsáních principech se budu snažit vycházet z výše uvedeného motto.

Na počítačové morfologii neboli automatické morfologické analýze (určování gramatických a lexikálních kategorií, resp. lemmatizaci libovolných slovních tvarů) a analýze (generování paradigmatických tvarů od základní podoby slova) se pro češtinu pracuje již mnoho desetiletí. Je dokonce možné sledovat jednotlivé články řetězu setrvalého úsilí o zdokonalování vnitřních datových formátů a algoritmů na straně jedné, z druhé strany potom projevy ctižádosti dosáhnout většího výkonu než systémy konkurenční, resp. evolučně předcházející. Oba dva akademické systémy, pražský a brněnský, tak mají své otce zakladatele a matky zakladatelky (SGALL 1960, 1967, KONEČNÁ – HRONEK 1962, OSOLSOBĚ 1996), množství trpělivých následovníků (WEISHEITLOVÁ 1981, WEISHEITLOVÁ – KRÁLÍKOVÁ – SGALL 1982, BÉMOVÁ – KRÁLÍKOVÁ 1988, HAJIČ 1994, 2004, 2010, SPOUSTA 2005, SEDLÁČEK 1999, 2004), solitéry vlastních cest (SKOUMALOVÁ 1997, GRÁC 2006), ale i více či méně razantní reformátory (KOLOVRATNÍK – PŘIKRYL 2008, HLAVÁČOVÁ 2009a, ŠMERK 2007b, 2008, 2009, 2010). Pokud k morfologické analýze připočítám i tzv. značkování (tagging), resp. disambiguaci neboli kontextové zjednoznačnění (OLIVA – HNÁTKOVÁ – PETKEVIČ – KVĚTOŇ 2000, HAJIČ – KRBEC – KVĚTOŇ – OLIVA – PETKEVIČ 2001, PETKEVIČ 2006, SPOUSTOVÁ 2007, ŠMERK 2007a, SPOUSTOVÁ – HAJIČ – VOTRUBEC – KRBEC – KVĚTOŇ 2007, SPOUSTOVÁ – HAJIČ – RAAB – SPOUSTA 2009), měly by být podle délky vývoje a počtu úprav dnešní podoby těchto nástrojů velmi vyspělé, snad téměř dokonalé. Poslední příspěvky k této problematice (OSOLSOBĚ 2006, 2007a, b, 2008a, b, KUČERA 2007, HLAVÁČOVÁ – OSOLSOBĚ 2008, HNÁTKOVÁ – JELÍNEK – PETKEVIČ 2009, HLAVÁČOVÁ 2009a, ŠMERK 2010) však budí trochu jiný dojem.

Pražský systém v nedávné době (alespoň návrhově) prošel rekonstrukcí v oblasti morfologických vzorů (nové kategorie pro rozlišování variant, adjektivní flexe má např. jeden sdílený vzor pro formálně blízká adjektiva i zájmena, více se objevuje dědičnost ve spojení s rozdílovými, tzv. kritickými kombinacemi), novinkou je i vícenásobné lemma pro juxtapozita typu *včera+s*, *na+ň* apod., inovován byl zejm. prefixový guesser atd. Největší potíže však

<sup>88</sup> V překladu B. Palka: „Jména nebud'ěž množována, pokud to není nutné.“ (Palek 1989: 93)

přetrvávají léta a zásadní změny v architektuře se již vzhledem k tradici (např. využití pro značkování korpusů ÚČNK) nezdají jako reálné. Nejslabším místem pražského přístupu je, podle mého soudu, koncepce lemmatu, a tedy i ohýbacích vzorů (srov. Šmerk 2010). Nepravidelné tvary jsou přidávány k těm odvozovaným pravidelně podle vzorů (i několika pro jedno slovo) tak, že jejich vzájemnou souvislost zajistí jediné zařazení pod společné lemma, které je do značné míry arbitrární. Uživatelé korpusů tak často podnikají dobrodružná hledání, pod jaké lemma autoři „schovali“ to nebo ono slovo či tvar. Brněnská koncepce se naopak (historicky) vydala cestou jednoho vzoru pro jedno slovo, byť by to měl být vzor neproduktivní, vytvořený pouze pro jednu nepravidelně se ohýbající lexikální jednotku. Lemma je v tomto přístupu libovolně vybraný tvar paradigmatu. Odvrácenou stranou elegantního řešení nepravidelností flexe je nepředstavitelně rozsáhlý systém přibližně 1800 vzorů, díky nimž je jakýkoli další rozvoj (např. přiřazování nových slov) velmi komplikovaný. Nejnověji se tak kromě úprav implementačního charakteru (jádro analyzátoru nyní představuje několik acyklických konečněstavových automatů) začíná prosazovat odklon od rozsáhlé soustavy vzorů směrem k aplikaci elementárních fonologických pravidel (např.  $d' + e = dě$  apod.), zásad dědičnosti a upřednostnění flektivních rysů před redundantními koncovkovými množinami.

Přes všechny naznačené vývojové impulzy přetrvává u obou systémů (v různé míře) problém s přegenerováváním (správné tvary plus některé navíc) a globální přístup k analýze i syntéze jako k práci s řetězcí (odeber, nahraď apod.), aniž by šlo o něco jiného než o výsledek bez ohledu na způsoby, které k němu vedou (tzv. technická řešení). Pražský i brněnský nástroj počítá s nějakou podobou odhadování vlastností neznámých slov (při společném zakončení nebo při obměně začátku slova), kromě toho existují na obou stranách různé přístupy k propojení flektivních a derivačních vztahů (PALA – SEDLÁČEK – VEBER 2004, PALA – HLAVÁČKOVÁ 2007, KLÍMOVÁ 1999, 2001, 2010, HLAVÁČKOVÁ – KLÍMOVÁ 2004). Ani jeden systém se však nedokáže vypořádat s analytickými tvary, podobně se nijak neřeší víceslovné lexikální jednotky. Ač je pro zařazení k flektivní třídě nezbytné přihlížet k výslovnostní podobě, morfologické nástroje pracují pouze v grafematickém režimu. Posledním závažným nedostatkem je budování strojových slovníků v naprosté většině nad vybranou teoretickou koncepcí a lexikálními zdroji (nejčastěji SSJČ), takže není vůbec jasné, o jaké modelování systému jazyka vlastně jde (moderní, lehce archaizující, synchronní?). Navzdory některým proklamacím (např. Osolsobě 1996) existují extenze systémů pouze do oblastí současné substandardní (zejm. obecněčeské), příp. pravidelné nářeční podoby češtiny. Jakékoli úspěšné pokusy o zpracování např. staročeského tvarosloví mi nejsou známy, stejně tak jako způsob, jakým užívané formalismy adaptovat pro potřeby morfematické analýzy, tedy překročit rámec členění slovních tvarů na část neměnnou (tzv. kmen) a variabilní (tzv. intersegmenty a/nebo koncovky).

Přesto se domnívám, že neexistuje jiná cesta, jak automaticky segmentovat česká slova na morfémy, než právě vhodné propojení s počítačovou morfologií. Vedou mě k tomu dva základní poznatky z dlouhodobého soužití s morfémy, a sice:

- 1) morfematické hranice se při ohýbání, tedy počítáme-li s většinovou koncepcí nulových morfémů, nemění
- 2) hranice zůstávají zachovány také při fonologické transkripci<sup>89</sup>, bez níž je např. začlenění některých slov cizího původu do systému české deklinace nemožné (srov. např. *Hortobágy* /-át', -ád'e .../, *Boulogne* /-oň, -ňe .../ apod.)

<sup>89</sup> Že nejde o zjištění nijak rozšířené, svědčí např. nedávný přípodotek V. Cvrčka: „*We have to admit that some of the problems caused by the difference between written and spoken representation could be solved by using phonologically transcribed texts. The result of such analysis would be interesting, but then we would not be able to assign exactly the morph boundary back to the written form. Furthermore, we still might want to know what the realisation of morphs looks like in the written language for the purpose of automatic corpus morphemic*“

Otázka tedy nezní proč, ale *jak*? Největší radostí (na straně spisovatelů a překladatelů) a naopak největší strastí (na straně učitelů češtiny jako cizího jazyka a tvůrců počítačových algoritmů) je neuvěřitelná variabilita a mnohotvárnost morfologie češtiny, která se promítá do nikdy nezodpověditelné otázky, kolik že skutečných flektivních vzorů pro češtinu potřebujeme. Připomínaný brněnský analyzátor jich vyžaduje přes 1800, „vícejazyční“ lexikografové se pohybují v intervalu od 213 (FRONEK 2005, 2007) do 708 (KOPECKÝ – FILIPEC – LEŠKA 1976). Moje zkušenosti s vytvářením (psané i mluvené) podoby morfematically organizovaných flektivních paradigm nad slovníkem E. Slavičkové říkají, že bez ohledu na doprovodné morfematically alternace je při zohlednění všemožných dublet, triplet a dalších variačních parametrů jen na úrovni (finálních) gramatických morfémů, tedy řekneme koncovek, (u sloves včetně konektémů) potřeba minimálně 617 různých kombinací. Na materiálu o velikosti cca 63 tisíc lemmat je produktivnost omezena horní hranicí 14 166 přiřazených slov, 428 „vzorů“ potom nesdílí více než pět slov.

Vývojové tendence české flexe, bohatě badatelsky zkoumané (např. HEBAL-JEZIERSKA 2006, 2008, TUŠKOVÁ 2006, HOLUBOVÁ 2006, BERMEL 2004, 2006, 2010B, ŠIMANDL 2010 aj.), však ukazují, že formální variabilitu nelze v rozumné míře vtěsnat do neměnných škatulek-vzorů, jak dokládá např. velmi hrubá aproximace přechodu mezi tradičními substantivními vzory píseň a kost v MČ2, kde je vymezeno 8 pracovních podvzorů. Budiž to příklad typu unus pro multis, takže zbývá jen odpovědět, jak v této věci postupovat jinak. Alternativní řešení si však také zaslouží úsilí o formální zpracování české slovo tvorby (srov. např. PALA – HLAVÁČKOVÁ 2007, KLÍMOVÁ 2010), které se dnes na jedné straně projevuje ve vytváření redundantních slovo tvorných vzorů pouhou subsumací vzorů tvaroslovných, na druhé straně potom vede k poněkud „donkichotské“ snaze vytvářet seznamy derivačních vztahů získaných četnými sondami do vygenerovaného inventáře morfologického slovníku. Můj návrh je, domnívám se, podstatně jednodušší, obecnější (neomezeně rozšiřitelný) a využitelný v širokém spektru možných aplikací (viz kap. 9). Základem je, jak ostatně jinak, morfematically segmentace slov.

## 8.2 Architektura navrhovaného systému

Základní myšlenkou návrhu je datová reprezentace, v níž by funkci základní stavební jednotky plnila taková lingvisticky adekvátní entita, která se v jazyce *a) vyskytuje dostatečně často* (má vysokou míru ekonomičnosti) – to nesplňují celá slova, *b) dá se většinou jednoznačně vymežit* – tomu odporuje např. slabika a *c) kombinuje se pokud možno omezeně a převážně předvídatelným způsobem* – z toho důvodu nepřipadá v úvahu foném. Jediným vhodným kandidátem je tak morfém, jednotka minimálního spojení formy a významu (funkce). Obecná představa počítá s existencí *různých typologických množin morfémů*, jejichž základními operacemi ***sjednocení***, ***průnik*** a ***rozdíl*** vznikají jak jednotlivá slova, jejich syntetické i analytické tvary, tak ustálená spojení (kolokace). Vyjadřovacími prostředky různorodých vztahů (vnitroparadigmatických i vněparadigmatických) jsou odkazy (adresy) typu M:N. Tento omezený repertoár stavebních kamenů je navíc otevřený jakýmkoli změnám<sup>90</sup>. Zařazení nového prvku znamená vybudování nových vztahů, pokud tím zároveň

---

*tagging, or for theoretical reasons. Thus, if we took phonologically transcribed texts and analysed them, we would have to transcribe back identified morphs into their written form, which might be almost impossible.*“ (CVRČEK 2009)

<sup>90</sup> Otázkami zásad náležitě morfematically segmentace v češtině se v této kapitoly již dále zabývat nebudu. V platnosti tak zůstávají kritéria uplatňovaná Slavičkovou i slovenskými morfematically, tedy sledování opakovatelnosti jednotek na paradigmatické a syntagmatické ose za současného přihlížení k sémantické stránce entit vyčleňovaných i motivujících. Kromě toho je vždy nutné reagovat na změny v systému, zde v strojovém slovníku, tak aby se motivovanost jednotek s nárůstem jeho velikosti spíše zvyšovala než ztrácela. Platí však také potřeba ověřovat hypotézy případných souvislostí ve výsledcích etymologických výzkumů, nikoli vlastním povědomím. Zvláště naléhavé je to u vlastních jmen.



vzrůstá motivovanost (pochoopení sémantických souvislostí), může nastat 1) *rozdělení jednoho původně celistvého morfému na dvě a více dílčích složek*, čímž se automaticky aktualizují všechny odkazy, které z jiných na tento morfém a z tohoto morfému na jiné vedou (tedy např.  $X \rightarrow Y$  na  $X_1 + X_2$  a  $Y_1 + Y_2$ ), 2) *spojení dvou původně samostatných morfémů v jeden* ( $X_1 \rightarrow Y_1$  a  $X_2 \rightarrow Y_2$  na  $X \rightarrow Y$ ). Problematika asociovaného, resp. rozštěpeného morfému (příp. cirkumfixu) je chápána v souladu se Slavičkovou jako realizace dvou a více synchronizovaných kroků (v jeden okamžik vstupují do jiné struktury a společně jsou eventuálně také redukovány, resp. modifikovány).

Anotační schéma předpokládá podobný, pochopitelně sofistikovanější systém popisu jednotek, jaký byl uveden v přechozí kapitole, i jejich **funkčního sřetězení** a **kombinací**. Formální stránka notace vychází z univerzálně modifikovatelného výměnného formátu XML: *zjednodušeně*:

morfém            <úvodní značka atribut<sub>1</sub>=““ atribut<sub>2</sub>=““ ...>**morfém**</ukončující značka>  
*neboli*            <m<sub>1</sub> a<sub>n</sub>></m<sub>1</sub>>

slovo              <úvod. zn. atr.1=““ ...><m<sub>1</sub> a<sub>n</sub>></m<sub>1</sub>> ... <m<sub>n</sub> a<sub>n</sub>></m<sub>n</sub>></ukonč. zn.>  
*neboli*            <s<sub>1</sub> a<sub>n</sub>></s<sub>1</sub>>

slovo<sub>syntet.</sub>    <s<sub>1</sub> a<sub>n</sub>></s<sub>1</sub>>

slovo<sub>analyt.</sub>    <s<sub>1</sub> a<sub>1</sub>=“1/2“ a<sub>n</sub>></s<sub>1</sub>> + odkaz<sub>1</sub> na <s<sub>2</sub> a<sub>1</sub>=“2/2“ a<sub>n</sub>></s<sub>2</sub>>

slovní spojení <úvod. zn. atr.1=““ ...> odkaz<sub>n</sub> na <s<sub>n</sub> a<sub>1</sub>=“n/n“ a<sub>n</sub>></ukonč. zn.>

V optimálním stavu systému platí, že každá jedinečná entita je uvedena pouze jednou a že jakékoli její další výskyty jsou řešeny pouhými odkazy na ni. Pro jednodušší představu je možné chápat odkaz jako adresu (systémové číslo), pod kterým je jednotka někde uložena. Symbolicky vyjádřeno by to vypadalo zhruba takto:

*zjednodušeně*:

množina prefixů     $P = \{m_1, m_2, m_3\} \sim \{1, 2, 3\}$

množina kořenů     $R = \{m_4, m_5, m_6\} \sim \{4, 5, 6\}$

množina interfixů     $I = \{m_7, m_8\} \sim \{7, 8\}$

množina sufixů     $S = \{m_9, m_{10}, m_{11}, m_{12}, m_{13}\} \sim \{9, 10, 11, 12, 13\}$

množina postfixů     $F = \{m_{14}, m_{15}\} \sim \{14, 15\}$

Potom by zápis konkrétního slova s<sub>1</sub> vypadal např. takto:

*zjednodušeně*:

<s<sub>1</sub> a<sub>n</sub>><m<sub>1</sub> a<sub>n</sub>></m<sub>1</sub>><m<sub>4</sub> a<sub>n</sub>></m<sub>4</sub>><m<sub>9</sub> a<sub>n</sub>></m<sub>9</sub>><m<sub>10</sub> a<sub>n</sub>></m<sub>10</sub>></s<sub>1</sub>>  
*nebo*    <s<sub>1</sub> a<sub>n</sub>>1 4 9 10</s<sub>1</sub>>

Za předpokladu, že slovo nemá žádné alternující morfémy, vypadá princip ohýbání zhruba takto (znak „|“ znamená disjunkci hodnot, mezi kterými stojí, a to i opakovaně v hranicích mezer):

<s<sub>1</sub> a<sub>n</sub>>1 4 9 10|10|10|10|10|10|11|10|12|11|10|10|12|13 </s<sub>1</sub>>

Pokud by 1 zastupovala prefix *pří*, 4 radix *tm*, 9 nefinální sufix *i* a 10 nulový finální sufix ( $\emptyset$ ), byla by výše uvedená formule předpisem deklinace substantiva *pří-tm-i- $\emptyset$* , tedy 11 pro *m*, 12 pro *ch* a 13 pro *mi* (prvních 7 alternativ pro singulár, zbývajících 7 pro plurál). Naprosto stejným způsobem lze zachytit jakoukoli variaci morfému na příslušné pozici. V případě, že by např. radix 5 znamenal *tem*, mohla být alternace kořene ve slově *tm-a* formálně zachycena takto 4|4|4|4|4|4|4|5|4|4|4|4|4 (*tem- $\emptyset$*  jako Gpl). Takový zápis by byl ale značně nepraktický, stačilo by uvést 4|5h, kde index *h* může zastupovat funkci Gpl. Zápis finálního gramatického morfu v předchozím příkladě však také není právě elegantní. Vzhledem k tomu, že se tak skloňuje velké množství slov, bylo by efektivnější definovat samostatnou množinu, která by obsahovala právě tuto kombinaci deklinačních morfémů. Tím se však oklikou vracím k soustavě vzorů, kterou jsem výše kritizoval. Paradox je to však jen zdánlivý, neboť schopnost rozdělit strukturu slova na funkční části alternující a nealternující (daná morfematickou segmentací) dovoluje vystačit si v případě substantiv s tradiční školskou soustavou 14 vzorů, ač může být (a často je) jejich identita různě zpochybňována (srov. např. ŠMERK 2010). Obecnou strukturu morfematických variací v kontextu navrhovaného způsobu zápisu je možné konkretizovat např. takto:

morfém <sub>1n</sub>	morfém <sub>2n</sub>	...	morfém <sub>nn</sub>
...	...	...	...
morfém <sub>1a</sub>	morfém <sub>2a</sub>	...	morfém <sub>na</sub>
<b>morfém<sub>1</sub></b>	<b>morfém<sub>2</sub></b>	...	<b>morfém<sub>n</sub></b>

Alternující morfémy s indexy *a* až *n* mohou představovat jakoukoli formální změnu na dané pozici (např. stylistické varianty *-ism-* a *-izm-*), nemusí jít nutně o alomorfy, resp. deklinační morfémy. Tímto způsobem je tak možné zachytit i např. supletivní morfémy a podobné nepravidelnosti.

Vrátím se ale ještě k problematice vzorů. Mnoho tradičních podvzorů (např. *les* a *jazyk* u neživotných maskulin podle dominantního vzoru *hrad*) se liší jen jedním, např. pádovým morfémem z celého souboru, resp. dubletou (Lsg *jazyku* a *jazyce*). Velmi výhodné je tak zavést princip dědičnosti vlastností, ovšem trochu jinak, než je dnes běžné (srov. HLAVÁČOVÁ 2009a, ŠMERK 2010). Postačí mi k tomu výše připomínané množinové operace. Pokud bych měl např. definovanou množinu morfémů pro flexi u vzoru *stavení*, tedy (plurál následuje bezprostředně za singulárem, oddělovačem je tečka)  $\emptyset.\emptyset.\emptyset.\emptyset.\emptyset.m.\emptyset.ch.m.\emptyset.ch.mi$ , bylo by možné v instrukci pro flexi uvádět jen rozdílné položky. Pro podvzor *jazyk* (oproti vzoru *hrad*:  $\emptyset.u.u.\emptyset.e.ě.em.y.ů.ům.y.y.ech.y$ ) tedy takto:  $.+a...-/ +u.-/ +e/ +u.....-| +ích$ . V Gsg se přidává *a*, ve Vsg odebírá původní hodnota (*e*) a naopak připojuje *u*, podobně v Lsg, a Lpl. Množinové operace průnik a rozdíl vytvoří z původního souboru nový, odpovídající původním požadavkům. V podobném duchu uvažuje např. Šmerk (2010). Pro mě jak však tento způsob zakládání nových entit na děděných vlastnostech jedním z pilířů návrhu, jeho užitečnost vidím daleko širěji. Stejným způsobem je totiž možné odvodit prakticky jakékoli rysy, např. obecně české tvary od spisovných (*+ama* apod.). Dosah je však ale hlubší, protože na stejném principu je možné vyjadřovat rozdíly mezi libovolnými dvěma prameny: např. tvary podle SSJČ, tvary doložené v korpusu apod.

Je pochopitelné, že naznačený postup nemůže pracovat pouze s formami, musí být k dispozici možnost podrobnější charakteristiky jednotlivých (flektivních) morfémů. To nabízí uvedená notace, u níž je možné definovat pro každý morfém libovolný počet atributů. Obligatorní může být rozlišení podle pořadí ve slově (prefix, sufix, radix atd.), vyznačení

funkce (gramatický, slovotvorný), zachycení hloubkové a povrchové funkce (např. původní konektém, nyní slovotvorný sufix apod.), především však určení pramene a údaj o podmínkách alternace. U slovesa tak může být vyznačena platnost alternant a dublet např. tímto způsobem:

*mhouriť*: (...) <R ... alt="vše" ...>mhouriť</R>|<R ... alt="imper.">mhuř</R> (...)

Ve všech tvarech se použije alomorf *mhouriť*, jen v imperativu vznikne dubleta *mhouriť-i/mhuř-Ø*. K odlišení platnosti stylistických, historických, teritoriálních a jakýchkoli jiných variant budou sloužit zase jiné atributy. Celkový počet morfémů se měnit nebude, a to ani v případě syntetického futura, kde lze v zápisu očekávat alternaci *Ø/po-*. Nespornou výhodou je také to, že určité atributy je možné přisuzovat jak jednotlivým morfémům, tak celému slovu. Tím se otevírá prostor pro rozlišení kategorií vyjadřovaných gramaticky i lexikálně (např. optimální × neoptimálnější). Týká se to především stupňování a životnosti (přirozeného rodu, tedy např. jako východisko pro tvary poses. adj.).

Naprosto stejný způsob organizace dat platí pro fonematickou reprezentaci. Příslušné identifikátory jednotlivých morfémů (systémová čísla, ID apod.) musejí být pro psanou a mluvenou podobu rozlišeny pouze jedním znakem, tak aby bylo možné kombinovat a srovnávat odpovídající si segmenty. Význam to má především pro analýzu vstupního textu, který může být psán neortograficky. Nekorektní tvar, např. *vjedomý*, se po transkripci do podoby */vjedomí/* již stává analyzovatelným pro fonologickou část systému. Důležité je načrtnout i předpokládaný způsob rozpoznávání zadaného slovního tvaru a s tím související techniku lematizace. U ní předpokládám podobný přístup jako v brněnském nástroji, kde může být za lemma označen jakýkoli tvar paradigmatu. V případě inkorporování různých variet jazyka je to zvláště důležité koncepční opatření. Při systematickém doplňování atributu pro pramen, resp. časové zařazení slovní zásoby je tak myslitelné uvažovat i o lematizaci podle konkrétního (nejčastěji asi slovníkového) zdroje.

Základem efektivní morfematické analýzy je systematicky zpracovaný slovník, v němž jsou u každého jednotlivého morfému vyznačeny všechny údaje o podmínkách distribuce variant, resp. odkazy na alternující morfémy<sup>91</sup>. Nezbytné je očíslování morfematických pozic od konce slova, velkou výhodou potom rozlišení slovnědruhové platnosti u všech segmentů, byť příslušnost ke konkrétnímu slovnímu druhu vzniká až na úrovni celého slovního tvaru. U analyzovaného slova se nejprve provede hledání mezi jednomorfémovými (některé předložky, adverbia apod.), v případě neúspěchu následuje předpoklad nulového finálního morfému. Seznam možných cest napříč slovníkem se tak omezí jen na taková hesla, která mohou končit morfémem bez materiální realizace. Není třeba podotýkat, že jsou zkoumány jen morfémy s vyznačenou pozicí první od konce. V případě úspěšně nalezených kandidátů se pokračuje v souladu se slovnědruhovou kompatibilitou a slovotvorným modelem (viz 7.2.8) na druhé pozici, z analyzovaného řetězce se nejdříve zkoušejí oddělit nejdelší možné morfémy z příslušného okruhu, teprve potom se prověřují hypotézy kratší. Identicky se po jednotlivých pozicích postupuje až do té doby, než je dosaženo shody v segmentaci s některým záznamem ve slovníku. Podle údajů uvedených v kap. 7.2.3 může být maximální počet morfémů ve slově 12, v korpusech jsou nejčastější slova třímorfémová. Analýza má tak v průměru kolem tří kol.

Pokud není nalezeno žádné vyhovující morfematické členění, vychází se z co nejdelší možné shody a zbývající část řetězce se porovnává s grafematickým, resp. fonematickým modelem pro jednotky o stejné délce a shodné slovnědruhové platnosti. Takový segment je pak označen za ten typ morfému (prefix, radix), jemuž se formálně nejvíc podobá, v každém

<sup>91</sup> Nastíněná optimalizace spočívající v nahrazení písmenného vyjádření číselnou hodnotou by měla být reverzibilní. Pro účely úprav strojového slovníku je pochopitelně výhodnější názornost, pro praktickou implementaci zase minimální datový objem.

případě musí mít libovolné slovo alespoň jeden slovní kořen. Je velmi pravděpodobné, že popsané odhadování neznámých segmentů nemůže být u nových slov příliš časté, většinou jde o aktualizace kombinující – podle běžných, dopředu známých pravidel – jednotky již ve slovníku obsažené. Proto by bylo na místě užívat pro tento postup spíše termínu analýza než hádání (guessing). Úspěšný morfologický analyzátor by měl umět poradit si také s rozpoznáváním analytických tvarů a víceslovných lexikálních jednotek. Jak jsem naznačil výše, jedním z atributů celého slova může být i odkaz na slovo jiné. Přenesení této informace do analyzovaného textu může výrazně napomoci procesu disambiguace. Pokud se ve vybraném úseku (klauzi) objeví slovní tvary opatřené příznakem „je členem většího spojení“ (nejspíše i kolikátým z kolika), stačí pouze vyhodnotit úplnost takového spojení a ostatní alternativy zavrhnout, v opačném případě ignorovat hypotézu víceslovnosti.

Naznačené obecné principy vnitřní organizace a fungování takového systému je nezbytně nutné doplnit o informace, jak uvést prozatím pouhý návrh do praxe. Domnívám se, že to nemusí být příliš obtížné. Jediné, co prozatím brání úpravě současné podoby paradigmatické nadstavby nad RMSČ do zamýšlené podoby, je z časových důvodů nakonec nerealizované rozlišení formálněmorfologických variant podle jednotlivých pramenů. Při přípravě dat jsem dal přednost konsenzuálnímu určování jednotlivých hodnot, ať již na úrovni slovnědruhové interpretace, především však při vyznačování vidu (z celkového počtu 14 601 sloves je 6 301 imperfektiv, 7 690 perfektiv a 610 biaspektuálních). Druhým problémem je struktura slovní zásoby RMSČ, která má v některých ohledech (nejen pravopisně) trochu archaizující charakter. Především však není dostatečně reprezentativní, a to zejm. ve vztahu k zastoupení některých slovních druhů.

Proto jsem přistoupil ke shromáždění lexikální zásoby relativně uzavřených slovních druhů (zájmen, prepozic, konjunkcí, partikulí a interjekcí) s důsledným uváděním pramenů, z nichž jsem čerpal (slovníků, gramatik, korpusů, ostatních analyzátorů a některých monografií). Základem pro budoucí systém je tak 285 zájmen, 762 prepozic (vč. vícesl.), více než 800 konjunkcí (vč. vícesl.), kolem 900 partikulí a přibližně 1 300 interjekcí. Doplnění další (pravidelněji tvořené) slovní zásoby a její začleňování do flektivního systému češtiny již může využívat některé postupy popsané výše. Revize hranic segmentací připadá v úvahu při zamýšleném rozlišování formálněmorfologických výjimek a variant ve vazbě na konkrétní zdroje. Nejvyšší prioritu má kvůli velkému aplikačnímu potenciálu především SSČ a slovníková část Pravidel českého pravopisu, teprve potom mohou být vytěžovány prameny další. Předpoklady pro vytvoření nového, morfematicky orientovaného analyzátoru a syntetizátoru češtiny proto pokládám za splněné.

### 8.3 Shrnutí

První část kapitoly kriticky hodnotí dva akademické systémy automatické počítačové morfologie, tzv. pražský a brněnský, druhá část potom představuje nejobecnější principy vlastního návrhu počítačové reprezentace morfematické struktury češtiny. Jeho základní devízou je integrace morfematické analýzy jako dílčí úlohy analýzy morfologické, což umožňuje optimalizovat některé tradiční přístupy k využití vzorů pro algoritmizaci české flexe. Východiskem proponovaného formalismu je duální, grafematické a fonematické zpracování jazyka. Představený způsob analýzy vstupního textu nabízí některé nové možnosti pro explorace slovotvorné a pro určování neznámých slov (tzv. guessing).

## 9 Význam jazykové univerzálie v globalizovaném světě (závěrem)

Téma morfemické struktury současné češtiny, jak z ní celý název této práce, bylo v předcházejících kapitolách probíráno z různých hledisek i z perspektiv mnoha různých oborů. Problematika vnitřní struktury slov má sice značně interdisciplinární povahu, je to však jen důsledek skutečnosti jiné, totiž že se bezprostředně dotýká podstaty jazyka jako takového, že jde pravděpodobně o nejdůležitější jazykovou univerzálii, kterou jsme schopni si uvědomit. Pokud je stále ještě předmětem komunikace směna zkušeností, názorů a postojů, neexistuje žádná efektivnější forma, než jakou zosobňuje jazykový kód. V pozadí jeho úspěchu potom stojí diskretnost jednotek a jejich mimořádný kombinatorický potenciál. Nejmenším prvkem jazyka, který umožňuje sdělování obsahů myslí, je morfém. Jakkoli se to může zdát neuvěřitelné, pouhých 100 nejčastějších morfémů (viz kap. 7.2.6.5–10) je v češtině zodpovědných za celých 61 % psaného textu. Podobná čísla budou pravděpodobně platit pro naprostou většinu světových jazyků. Čím se naopak jednotlivé jazyky budou bezesporu lišit, je míra uvědomění si tohoto fenoménu, která se odráží v intenzitě zájmu o tuto oblast jazyka. Podnětů, které mohou atraktivitu morfematiky – jako dotčené lingvistické disciplíny – v očích nejširší veřejnosti, ale především také vývojářů a programátorů – zodpovědných za proteovské parametry informační společnosti – zvýšit, je bezpočet. Pokusím se připomenout jen některé z nich.

V každodenním životě se všichni běžně setkáváme s technologiemi, v nichž se skrývají systémy pro analýzu a syntézu řeči, z nichž nejsložitější jsou ty, které mají v reálném čase převádět akustický signál do psané podoby. Týká se to např. titulovacích zařízení určených především pro hluchoněmé televizní diváky, prepisu nejrozličnějších mediálních relací zajišťovaných pro zpravodajské a rešeršní účely nebo diktovacích souprav využívaných soudními úředníky a také nevidomými pro ovládání počítače a další techniky. V hloubi těchto neobyčejně složitých hlasových systémů jsou integrovány slovníky nejčastějších slov, které pomáhají rozpoznat (segmentovat) nejednoznačné sekvence spojitě řeči (srov. např. PSUTKA ET AL. 2006). Pro flektivně chudou angličtinu je obvyklý rozsah takových slovníků zvíci maximálně několik set tisíc slov plně dostačující, v češtině je však nezbytné obsadit tuto kapacitu pouze nejčastějšími slovními tvary (z odhadovaného množství více než 10 milionů), nikoli lemmaty. Výsledky potom nejsou takové, jaké by si uživatelé přáli mít. Právě zde by tak morfemická segmentace mohla výrazně přispět, neboť by umožnila úspěšněji využít paměťový prostor a zvýšit také rychlost zpracování. Potvrzují to např. aplikace pro ruštinu (RONZHIN – KARPOV 2004).

Oblastí, která se týká ještě většího množství uživatelů, je automatická kontrola pravopisu (a gramatiky), jež bývá integrována do velkého množství textových editorů a aplikačních rozhraní pro e-mailovou komunikaci. Častým jevem rutinní práce s těmito nástroji je situace, kdy se v textu objevuje tolik neznámých slov, že se zkrátka vnitřní zásobník přeplní a upozorňování na potenciální chyby, např. překlepy přestane pro tento dokument fungovat. Problém nespočívá v tom, že se funkce deaktivovala, ale v množství korektně napsaných slov, jež program nezná. Také v tomto případě by výrazně prospěla dovednost odhalovat vnitřní strukturu slov. Naprostá většina nerozpoznaných slov (pokud nejde skutečně o omyl) se totiž dá „sestavit“ z morfémů obsažených ve slovech (tvarech), která program zná a verifikuje. Právě tak by jistě bylo užitečné mít po ruce rádce, který rozezná, jestli je stylistická úroveň textu konzistentní, zdali komunikát nepřekračuje slohovou úroveň směrem níže, avšak také vzhůru. Morfemická delimitace nabízí prostor pro distribuci jednou zadané informace na všechna dotčená (derivovaná) místa (např. expresivní kořen zjevně ovlivní stylistickou příznakovost všech svých odvozenin apod.).

Vývoj internetu je nezadržitelný a touha uživatelů prodrat se houštinou bezcenných informací k těm, které je skutečně zajímavé, zaměstnává mnoho vývojových týmů po celém

světě. Ctižádost programátorů dobrat se automatickým způsobem prostřednictvím formy textů také jejich významů je dnes velmi běžná. Mluví se např. o tzv. sémantickém webu, slovních sítích, ontologiích, expertních systémech, strojovém překladu apod. (srov. MANNING – SCHÜTZE 1999). V odvětví informatiky zabývajícím se tzv. dolováním dat se užívají různé techniky pro automatickou tematickou klasifikaci textů, nejběžnější je princip označovaný jako skrytá sémantická analýza nebo skryté sémantické indexování. Podstata spočívá ve vhodném rozvržení slovní zásoby do pojmových tříd (pro češtinu srov. KLÉGR 2007) a v označování slov podle příslušnosti do jednotlivých taxonomických skupin. Výsledný průměr (vektor) sémantických odstínů konkrétních slov specifikuje téma celého textu. I zde by mohla být morfemická segmentace velkým přínosem, a to zejména při výchozím rozdělení slov do pojmových kategorií. Ve většině případů by totiž zřejmě stačilo charakterizovat pouze množinu kořenů.

Příklady, kde by schopnost členit slova na jednotlivé funkční segmenty mohla najít své další uplatnění, je však mnohem více. Dalo by se tak ještě mluvit např. o experimentálním psycholingvistickém výzkumu (srov. SMOLÍK 2009), o lingvistické větvi strojového překladu nebo o učení cizinců češtině. Mnohem důležitější je však nezapomenout, že proslulý furor orthographicus řadí na českých školách vlastně neprávem, protože jen málokdy dostanou studenti dostatek informací o souvztažnosti psané a mluvené podoby, stejně tak jako o důležitosti rozlišovat psaní určitých hlásek a jejich skupin podle toho, v jaké části slova se vyskytují. V tom všem jim může morfemická segmentace a identifikace výrazně prospět. Neméně by jim však prospělo, kdyby se dozvěděli něco o historii zkoumání vnitřní struktury českých slov, která sahá přinejmenším až do časů Komenského, nebo kdyby alepoň jedinkrát slyšeli o tom, kdo byla a stále ještě je Eleonora Slavičková a čemu se ve své lingvistické kariéře věnovala.

## Resumé

Morfematický výzkum češtiny zažil největší rozmach v šedesátých a sedmdesátých letech 20. století. Od vydání Retrográdního morfematického slovníku češtiny (1975) Eleonory Slavičkové a Komárkových Příspěvků k české morfologii (1978) badatelský zájem o tuto oblast lingvistiky postupně upadá. Pokus o obhajobu morfematiky jako samostatné disciplíny, zhodnocení dosavadních teoretických i empirických poznatků a odůvodnění volby formálního počítačového zpracování jako jediné možné další perspektivy je leitmotivem všech devíti kapitol monografie. Její interdisciplinární charakter se projevuje v hledání podnětů pro návrh algoritmického pojetí morfematické analýzy a syntézy, jímž práce vrcholí, např. ve sféře kognitivních věd nebo obecné sémiotiky. Poznání základních principů fungování mentálního slovníku včetně podstaty teorie aktivovanosti jazykových jednotek ukazuje, že počítačový přístup k tradičním lingvistickým tématům a metodám se může inspirovat nejen v teoretických oborech. Za tradiční počátek ústředního termínu, kterým je systémově morfém realizující se v užití jazyka jako morf, se pokládá jeho zavedení Janem Baudouinem de Courtenay v osmdesátých letech 19. století. Na příkladu velkých osobností vývoje české lingvistiky, J. A. Komenském, V. J. Rosovi, J. Dobrovském a F. L. Čelakovském, je ukázáno, že pojem morfému má v češtině mnohem delší trvání, než se obvykle soudí. Velmi podrobně jsou diskutovány zásady morfematické segmentace, tedy výchozí princip opakovatelnosti, využívaný i v informatice, doplňovaný sémantickým kritériem. Velký prostor je věnován kvantitativnímu představení důležitých charakteristik morfematické stavby současné češtiny, jejichž výchozí datovou základnou se stal mnohostranně anotovaný slovník E. Slavičkové, zpracovávaný v původní grafematické, ale nově též v transkribované fonematické podobě. Hodnoty týkající se produktivity systému jsou konfrontovány s frekvenčními údaji z velkých synchronních korpusů češtiny o celkovém úhrnu 1,2 miliardy textových slov. Díky učiněnému zjištění, že počet morfémů a jejich hranice se při flexi nemění, bylo možné určit, jak velká část korpusové evidence je morfematicky členitelná a identifikovatelná pouze s využitím dosavadních výsledků. Týká se to přibližně 93 % obsahu. Proponovaný systém morfematické analýzy a syntézy zúročuje jak zkušenosti tvůrců dvou hlavních morfologických analyzátorů pro češtinu, tak i detailní rozbor českého morfematického systému. Naznačeny jsou např. způsoby, jak se vyrovnávat s problematikou analytických tvarů, rozpoznávání neznámých slov nebo závislosti formalizace na konkrétním popisu jazyka.

## Seznam klíčových slov

morf  
morfém  
alomorf  
morfematika  
morfotaktika  
morfonologie  
formální morfologie  
kvantitativní charakteristiky  
automatické počítačové zpracování  
morfematické slovníky  
jazykové korpusy

## Summary

The morphemic research of the Czech language witnessed its largest boom in the 1960s and 1970s. Since the appearance of *Retrográdní morfemický slovník češtiny* (1975) by Eleonora Slavičková and of Komárek's *Příspěvky k české morfologii* (1978), the interest of researchers in this area of linguistics has been waning. The leitmotifs of all nine chapters of this monograph are an attempt to defend morphemics as a stand-alone discipline, an evaluation of the theoretic and empirical knowledge gathered so far, and the justification of formal computer processing as the only promising approach for the future. The interdisciplinary character of the present work manifests itself in the search for impulses for proposing an algorithmic approach to morphemic analysis and synthesis – which is the culmination of the monograph – e.g. in cognitive sciences and general semiotics. An examination of the main principles of how the mental lexicon works, including the theory of activation of language units, shows that a computational approach to traditional linguistic topics and methods can borrow inspiration not only from theoretical fields. The central term, namely morpheme, which is realized as morph in the language usage, is widely believed to have been introduced by Jan Baudouin de Courtenay in the 1880s. Using the great personalities of the Czech linguistics such as J. A. Komenský, V. J. Rosa, J. Dobrovský and F. L. Čelakovský as examples, the present work shows that the concept of morpheme has a longer tradition in Czech than is generally thought. The principles of morphemic segmentation are discussed in much detail, specifically the initial principle of reproducibility, which is also used in computer science, complemented by a semantic criterion. Much attention is paid to the quantitative presentation of important characteristics pertaining to morphemic structure of contemporary Czech. The source of data is the well annotated dictionary of E. Slavičková, which is elaborated in its original graphematic form but newly also in a transcribed phonemic form. Values regarding productivity of the system are compared with frequency data from large, synchronous corpora of the Czech language containing a total of 1.2 billion text words. Thanks to the discovery – achieved in the present work – that the number of morphemes and their borders do not change due to inflexion, it was possible to determine how big a part of the corpus evidence can be classified morphemically and identified only using the existing results: 93 percent. The proposed system of morphemic analysis and synthesis leverages not only the experience of the authors of the two main morphologic analysers for Czech but also a detailed analysis of the Czech morphemic system. For example, the present work suggest ways to solve the problem of analytic forms, recognition of unknown words or the dependence of formalisation on a specific description of the language.

## Key words

morph  
morpheme  
alomorph  
morphemics  
morphotactics  
morphonology  
formal morphology  
quantitative characteristics  
automatic computational processing  
morphemic dictionaries  
language corpora



## Literatura

- ADAM, R. (2006–2007). O nejvlastnější mateřštině. *Český jazyk a literatura*, 2006–2007, roč. 57, č. 5, s. 244–246. ISSN 0009-0786.
- ADAM, R. (2007). „Senátorka umřela.“: K jazykové kultuře veřejnoprávního televizního a rozhlasového zpravodajství. In Uličný, O. (ed.). *Euro litteraria & eurolingua 2006*. Liberec: Technická univerzita v Liberci, 2007, s. 108–115. ISBN 80-7372-040-X.
- ADAM, R. (2009). Za empiricky adekvátní teorii písemné komunikace. *Slovo a slovesnost*, 2009, roč. 70, č. 3, s. 221–230. ISSN 0037-7031.
- ALTMANN, G. T. M. (2005). *Výstup na Babylonskou věž: Otázky jazyka, mysli a porozumění*. Praha: Triáda, 2005. ISBN 80-86138-70-4.
- ANDERSON, S. R. (1992). *A-morphous morphology*. Cambridge: Cambridge Univ., 1992. ISBN 0-521-37260-7.
- ARONOFF, M. (1979). *Word formation generative grammar*. 2. print. Cambridge: MIT Press, 1979. ISBN 0-262-510017-0.
- ARONOFF, M. (1994). *Morphology by itself*. Cambridge, Mass.: MIT Press, c1994. ISBN 0-262-51072-3.
- ARONOFF, M. – VOLPE, M. (2006). Morpheme. In Brown, K., et al. *Encyclopedia of Language and Linguistics*. 2nd edition. Amsterdam; Boston; Heidelberg; London; New York; Oxford; Paris; San Diego; San Francisco; Singapore; Sydney; Tokyo: Elsevier, 2006, s. 274–276. ISBN 0-08-044299-4.
- AUGST, G. (1975). *Lexikon zur Wortbildung: Morpheminventar*. Tübingen: TBL Gunter Narr, 1975. Band 24.1 [A–G], 24.2. [H–R] a 24.3 [S–Z]. 1306 s. ISBN 3-87808-624-5. dostupné na: <<http://books.google.com/books?id=mDnQxw7aapAC&hl=cs>> [cit. 9. 8. 2010].
- BARTHES, R. (1967). *Nulový stupeň rukopisu; Základy sémiologie*. Praha: Československý spisovatel, 1967.
- BEARD, R. (1995). *Lexeme-morpheme base morphology: A general theory of inflection and word formation*. Albany: State University of New York Press, c1995. ISBN 0-7914-2472-3.
- BEDNAŘÍKOVÁ, B. (2009). *Slovo a jeho konverze*. Olomouc: FF UP, 2009. ISBN 978-80-244-2220-6.
- BEDNAŘÍKOVÁ, B. (2010). Slovo(tvorba) a text. In Čmejrková, S. – Hoffmannová, J. – Havlová, E. (eds). *Užívání a prožívání jazyka: K 90. narozeninám Františka Daneše*. Praha: Karolinum, 2010, s. 135–140. ISBN 978-80-246-1756-5.
- BÉMOVÁ, A. – KRÁLÍKOVÁ, K. (1988). K otázkám automatického zpracování českého tvarosloví. *Slovo a slovesnost*, 1988, roč. 49, č. 4, s. 285–295. ISSN 0037-7031.
- BERGENHOLTZ, H. (1976). *Zur Morphologie deutscher Substantive, Verben und Adjektive: Probleme der Morphe, Morpheme und ihrer Beziehung zu den Wortarten*. Bonn: F. Dümmler, 1976.
- BERKA, P. (2003). *Dobývání znalostí z databází*. Praha: Academia, 2003. ISBN 80-200-1062-9.
- BERMEL, N. (2004). V korpuse nebo v korpusu? Co nám řekne (a neřekne) ČNK o morfologické variaci v tvarech lokálu. In Hladká, Z. – Karlík, P. (eds.). *Čeština – univerzálie a specifika 5*. Praha: Nakladatelství Lidové noviny, 2004, s. 163–171. ISBN 80-7106-659-1.

- BERMEL, N. (2006). Tvary minulého přičestí v ČNK: táh, táhl, či táhnul?. In Čermák, F. – Blatná, R. (eds.). *Korpusová lingvistika: Stav a modelové přístupy*. Praha: Nakladatelství Lidové noviny. s. 19–34. ISBN 80-7106-865-9.
- BERMEL, N. (2010a). O tzv. české diglosii v současném světě. *Slovo a slovesnost*, 2010, roč. 71, č. 1, s. 5–30. ISSN 0037-7031.
- BERMEL, N. (2010b). Variace a frekvence variant na příkladu tvrdých neživotných maskulin. In Čmejrková, S. – Hoffmannová, J. – Havlová, E. (eds.). *Užívání a prožívání jazyka: K 90. narozeninám Františka Daneše*. Praha: Karolinum, 2010, s. 147–151. ISBN 978-80-246-1756-5.
- BIRD, N. (1982). *The distribution of Indo-European root morphemes*. Wiesbaden: Otto Harrassowitz, 1982. ISBN 3-447-02284-1.
- BLOOMFIELD, L. (1933/1965). *Language*. Repr. London: G. Allen and Unwin, 1965.
- BOJAR, O. (2008). *Exploiting linguistic data in machine translation*. PhD thesis at MFF UK, Institute of formal and applied linguistics. Prague, 2008.
- BORDAG, S. (2005). Unsupervised Knowledge-Free Morpheme Boundary Detection. In *Advances in Multilingual and Multimodal Information Retrieval*. Berlin: Springer, 2008, s. 881–891. ISBN 978-3-540-85759-4. Dostupné na: <<http://wortschatz.uni-leipzig.de/~sbordag/papers/BordagMorph05.pdf>> [cit. 9. 8. 2010].
- BORDAG, S. (2008). Unsupervised and Knowledge-free Morpheme Segmentation and Analysis. In *Advances in Multilingual and Multimodal Information Retrieval*. Berlin: Springer, 2008, s. 881–891. ISBN 978-3-540-85759-4.
- BOSÁK, J. – BUZÁSSYOVÁ, K. (1985). *Východiská morfémové analýzy: Morfematika, slovotvorba*. Bratislava: Veda, 1985.
- BOZDĚCHOVÁ, I. (1994). *Tvoření slov skládáním*. Praha: Institut sociálních vztahů nakladatelství, [1994]. ISBN 80-85866-04-08.
- BROWN, K., et al. (2006). *Encyclopedia of Language and Linguistics*. 2nd edition. Amsterdam; Boston; Heidelberg; London; New York; Oxford; Paris; San Diego; San Francisco; Singapore; Sydney; Tokyo: Elsevier, 2006. ISBN 0-08-044299-4.
- COATES, R. Morphophonemics. In Brown, K., et al. *Encyclopedia of Language and Linguistics*. 2nd edition. Amsterdam; Boston; Heidelberg; London; New York; Oxford; Paris; San Diego; San Francisco; Singapore; Sydney; Tokyo: Elsevier, 2006, s. 318–331. ISBN 0-08-044299-4.
- CREUTZ, M. (2003). Unsupervised Segmentation of Words Using Prior Distributions of Morph Length and Frequency. In *Annual Meeting of the ACL: Proceedings of the 41st Annual Meeting on Association for Computational Linguistic*. Vol. 1. 2003, s. 280–287. Dostupné na: <<http://lib.tkk.fi/Diss/2006/isbn9512282119/article2.pdf>> [cit. 9. 8. 2010].
- CREUTZ, M. (2006). *Induction of the morphology of natural language: Unsupervised morpheme segmentation with application to automatic speech recognition*. Helsinki, 2006. Dissertations at Helsinki University of Technology <<http://lib.tkk.fi/Diss/2006/isbn9512282119/isbn9512282119.pdf>> [cit. 9. 8. 2010].
- CREUTZ, M. – LAGUS, K. (2002). Unsupervised Discovery of Morphemes. In *Morphological and Phonological Learning: Proceedings of the 6th Workshop of the ACL Special Interest Group in Computational Phonology (SIGPHON)*. 2002, s. 21–30. Dostupné na: <<http://www.aclweb.org/anthology/W/W02/W02-0603.pdf>> [cit. 9. 8. 2010].
- CREUTZ, M. – LAGUS, K. – LINDÉN, K. – VIRPIOJA, S. (2005). Morfessor and Hutmegs: Unsupervised Morpheme Segmentation for Highly-Inflecting and Compounding Languages.

- In *Second Baltic Conference on Human Language Technologies*. Talinn, 2005. Dostupné na: <<http://www.cis.hut.fi/mcreutz/papers/Creutz05hlt.pdf>> [cit. 9. 8. 2010].
- CREUTZ, M. – LINDÉN, K. (2004). Morpheme Segmentation Gold Standards for Finnish and English. In *ACM Transactions on Speech and Language Processing (TSLP)*. 2007, vol. 4, issue 1, s. 1–33. ISSN 1550-4875. Dostupné na: <<http://www.cis.hut.fi/mcreutz/papers/Creutz04tr.pdf>> [cit. 9. 8. 2010].
- CVRČEK, V. (2008). *Regulace jazyka a Koncept minimální intervence*. Praha: Nakladatelství Lidové noviny, 2008. ISBN 978-80-7106-600-2.
- CVRČEK, V. (2009). Corpus-driven morphematic analysis. In Mahlberg, M. – González-Díaz, V. – Smith, C. (eds.). *Proceedings of the Corpus linguistics conference*. Liverpool: University of Liverpool, 2009. Dostupné na: <[http://ucrel.lancs.ac.uk/publications/cl2009/156\\_FullPaper.doc](http://ucrel.lancs.ac.uk/publications/cl2009/156_FullPaper.doc)>. [cit. 9. 8. 2010].
- CVRČEK, V., et al. (2010). *Mluvnice současné češtiny (I)*. Praha: Karolinum, 2010. ISBN 978-80-246-1743-5.
- CVRČEK, V. – VYBÍRAL, P. (2005). Mluvme, jak nám zobák narost. *Lidové noviny*, 13. 8. 2005, Příloha Orientace, s. 5. ISSN 0862–5921.
- CZÉREOVÁ, B. (2006). Vokální alternácie v radixoch. In Sokolová, M. – Ivanová, M. – Ološtiak, M. (eds.). *Morfematický výskum slovenčiny: Možnosti jeho štatistického elektronického a didaktického spracovanie*. Prešov: Filozofická fakulta Prešovskej univerzity v Prešove, 2006, s. 145–164. ISBN 80-8068-544-4.
- ČERMÁK, F. (1990). *Syntagmatika a paradigmatica českého slova II: Morfologie a tvoření slov*. Praha: Univerzita Karlova, 1990.
- ČERMÁK, F. (1995). Jazykový korpus: Prostředek a zdroj poznání. *Slovo a slovesnost*, 1995, roč. 56, č. 2, s. 119–140. ISSN 0037-7031. Dostupné na: <<http://is.muni.cz/el/1421/jaro2008/CJBB85/um/Cermak.txt>>
- ČERMÁK, F. (2001). *Jazyk a jazykověda: Přehled a slovníky*. 3., dopl. vyd. Praha: Karolinum, 2001. ISBN 80-246-0154-0.
- ČERMÁK, F. (ed.). (2007). *Slovník Karla Čapka*. Praha: Nakladatelství Lidové noviny, 2007. ISBN 978-80-7106-915-7.
- ČERMÁK, F. (2008a). Diskrétní jednotky v jazyce: případ cirkumfixů. *Slovo a slovesnost*, 2008, roč. 69, č. 1–2, s. 78–98. ISSN 0037-7031.
- ČERMÁK, F. (2008b). The Case of The Czech National Corpus: Its Design and Development. 2008. Dostupné na: <[http://www.clarin.lv/materiali/Cermak\\_Riga-en.doc](http://www.clarin.lv/materiali/Cermak_Riga-en.doc)> [cit. 9. 8. 2010].
- ČERMÁK, F. (2010). *Lexikon a sémantika*. Praha: Nakladatelství Lidové noviny, 2010. ISBN 978-80-7422-020-3.
- ČERMÁK, F. – CVRČEK, V. (eds.). (2009). *Slovník Bohumila Hrabala*. Praha: Nakladatelství Lidové noviny, 2009. ISBN 978-80-7106-488-6.
- ČERNÝ, J. (1996). *Dějiny lingvistiky*. Olomouc: Votobia, 1996. ISBN 80-85885-96-4.
- ČERNÝ, J. – HOLEŠ, J. (2004). *Sémiotika*. Praha: Portál, 2004. ISBN 80-7178-832-5.
- DANEŠ, F. (2000). On Linguistic Strata (Levels). In Daneš, F. *Jazyk a text II*. Praha: FF UK, 2000, s. 13–30. ISBN 80-85899-86-8
- DANEŠ, F. – DOKULIL, M. – KUCHAR, J., et al. (1967). *Tvoření slov v češtině 2: Odvozování podstatných jmen*. Praha: Academia, 1967.
- DASGUPTA, S. – NG, V. (2007). High-Performance, Language-Independent Morphological Segmentation. In *Proceedings of Human Language Technology (NAACL)*. Rochester (New

- York), 2007. Dostupné na: <<http://www.hlt.utdallas.edu/~vince/papers/naacl07.pdf>> [cit. 9. 8. 2010].
- DÉJEAN, H. (1998). Morphemes as Necessary Concept for Structures Discovery from Untagged Corpora. In *ACL Workshops: Proceedings of the Joint Conferences on New Methods in Language Processing and Computational Natural Language Learning*. Sydney, 1998, s. 295–298. ISBN 0-7258-0634-6. Dostupné na: <<http://acl.ldc.upenn.edu/W/W98/W98-1239.pdf>> [cit. 9. 8. 2010].
- DOBROVSKÝ, J. (1953). Tvořivost slovanské řeči (1799). In Jedlička, B. (ed.). *Josef Dobrovský: Výbor z díla*. Praha: SNKLHU, 1953, s. 289–299.
- DOBROVSKÝ, J. (1799). *Die Bildsamkeit der Slawischen Sprache an der Bildung der Substantive aund Adjective in der Böhmischen Sprache dargestellt*. Prag, 1799.
- DOKULIL, M. (1962). *Tvoření slov v češtině I: Teorie odvozování slov*. Praha: Nakladatelství Československé akademie věd, 1962.
- DOKULIL, M. (1997). Status a funkce tzv. kmenotvorné přípony slovesné v slovanském tvoření slov (v pohledu diachronním a synchronním). In Dokulil, M. *Obsah – výraz – význam I*. Praha: FF UK, 1997, s. 191–200. ISBN 80-85899-37-X.
- DOKULIL, M. – KUCHAR, J. (1977). Slovtvorná charakteristika cizích slov. *Naše řeč*, 1977, roč. 60, č. 4, s. 169–185. ISSN 0027-8203.
- DOUBRAVOVÁ, J. (2002). *Sémiotika v teorii a praxi*. Praha: Portál, 2002. ISBN 80-7178-566-0.
- ĐUROVIČ, L. (1965). Ku klasifikácii morféme. *Slovo a slovesnost*, 1965, roč. 26, č. 2, s. 126–129. ISSN 0037-7031.
- ERHART, A. (1973). *Úvod do obecné a srovnávací jazykovědy*. Praha: SPN, 1973.
- ERHART, A. (2001). *Úvod do jazykovědy*. Brno: MU, 2001. ISBN 80-210-2669-3.
- [ESČ] KARLÍK, P. – NEKULA, M. – PLESKALOVÁ, J. (eds.). (2002). *Encyklopedický slovník češtiny*. Praha: Nakladatelství Lidové noviny, 2002. ISBN 80-7106-484-X.
- FLAJŠHANS, V. (1929). Josef Dobrovský a české tvoření slov. In Horák, J. – Murko, M. – Weingart, M. (eds.). *Josef Dobrovský 1753–1829: sborník statí k stému výročí smrti Josefa Dobrovského*. Praha, 1929, s. 81–88.
- FRONEK, J. (2005). Gramatické tabulky. In Máčelová-Van den Broecke, W. – Spěváková, D. *Česko-nizozemský slovník*. Voznice: Leda, 2005. ISBN 80-7335-057-2.
- FRONEK, J. (2007). *Velký anglicko-český, česko-anglický slovník*. Voznice: Leda, 2007. ISBN 978-80-7335-114-4. jen paradigmatika
- GERD, A. S. (2004). *Morfemika*. Sankt-Peterburg: Izdatel'stvo S.-Peterburgskogo universiteta, 2004. Netransl. název: Морфемика. ISBN 5-288-03484.
- GLEASON, H. A. (1969). *An introduction to descriptive linguistics*. Revised edition. London: Holt, Rinehart and Winston, 1969.
- GOLDSMITH, J. (2001). Unsupervised Learning of the Morphology of a Natural Language. *Computational Linguistics*, 2001, vol. 27, issue 2, s. 153–198. ISSN 0891-2017. Dostupné na: <<http://acl.ldc.upenn.edu/J/J01/J01-2001.pdf>> [cit. 9. 8. 2010].
- GRÁC, M. (2006). Tvorba morfologické databáze z neoznačkováného korpusu. Brno, 2006. Diplomová práce na FI MU. Dostupné na: <[http://is.muni.cz/th/50728/fi\\_m/Diplomova\\_praca.pdf](http://is.muni.cz/th/50728/fi_m/Diplomova_praca.pdf)> [cit. 9. 8. 2010].
- GREENBERG, J. H. (1957). *Essays in Linguistics*. Chicago: The University of Chicago Press, 1957.



- HAIJČ, J. (1994). *Unification Morphology Grammar: Software System for Multilanguage Morphological Analysis*. Prague, 1994. PhD thesis at MFF UK, Institute of formal and applied linguistics.
- HAIJČ, J. (2004). *Disambiguation of rich inflection: Computational morphology of Czech*. Praha: Karolinum, 2004. ISBN 80-246-0282-2.
- HAIJČ, J. ([2010]). Popis morfologických značek – poziční systém. Dostupné na: <[http://ucnk.ff.cuni.cz/doc/popis\\_znacek.pdf](http://ucnk.ff.cuni.cz/doc/popis_znacek.pdf)> [cit. 9. 8. 2010].
- HAIJČ, J. – KRBEC, P. – KVĚTOŇ, P. – OLIVA, K. – PETKEVIČ, V. (2001). Serial Combination of Rules and Statistics: A Case Study in Czech Tagging. In *Annual Meeting of the ACL: Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*. Toulouse, 2001, s. 268–275. Dostupné na: <<http://www.aclweb.org/anthology-new/P/P01/P01-1035.pdf>> [cit. 9. 8. 2010].
- HAIJČOVÁ, E. – PANEVOVÁ, J. – SGALL, P. (2002). *Úvod do teoretické a počítačové lingvistiky I: Teoretická lingvistika*. Praha: Karolinum, 2002. ISBN 80-246-0470-1.
- HAMADOU, A. B. (1986). A compression technique for arabic dictionaries: The affix analysis. In *International Conference On Computational Linguistics: Proceedings of the 11th conference on Computational linguistics*. Bonn, 1986, s. 286–288. Dostupné na: <<http://www.aclweb.org/anthology/C/C86/C86-1068.pdf>> [cit. 9. 8. 2010].
- HAMMARSTRÖM, H. (2006a). A Naive Theory of Affixation and an Algorithm for Extraction. In *ACL Workshops: Proceedings of the Eighth Meeting of the ACL Special Interest Group on Computational Phonology and Morphology*. New York, 2006, s. 79–88. Dostupné na: <<http://www.cs.chalmers.se/~harald2/sigphon06.pdf>> [cit. 9. 8. 2010].
- HAMMARSTRÖM, H. (2006b). Poor Man's Stemming: Unsupervised Recognition of Same-Stem Words. In *Information Retrieval Technology*. Berlin: Springer, 2006, s. 323–337. ISBN 978-3-540-45780-0.
- HARRIS, Z. S. (1960). *Structural linguistics*. 4th Impression. Chicago: The University of Chicago Press, 1960.
- HAUSER, P. (1959). Dobrovského práce o tvoření slov a domácí mluvnická tradice. In Jedlička, A. – Dvořák, K. (eds.). *Studie o jazyce a literatuře národního obrození*. Praha: SPN, 1959, s. 25–54.
- HAY, J. Morphotactics. In Brown, K., et al. *Encyclopedia of Language and Linguistics*. 2nd edition. Amsterdam; Boston; Heidelberg; London; New York; Oxford; Paris; San Diego; San Francisco; Singapore; Sydney; Tokyo: Elsevier, 2006, s. 335–337. ISBN 0-08-044299-4.
- HEBAL-JEZIERSKA, M. (2006). Možnosti a meze variantního nominativu plurálu měkkých typů životných maskulin ve světle Českého národního korpusu. In Štícha, F. (ed.). *Možnosti a meze české gramatiky*. Praha: Academia, 2006, s. 88–98. ISBN 80-200-1463-2.
- HEBAL-JEZIERSKA, M. (2008). Variantnost některých podstatných jmen v mluveném korpusu. In Kopřivová, M. – Waclawíčková, M. (eds.). *Čeština v mluveném korpusu*. Praha: Nakladatelství Lidové noviny, 2008, s. 89–96. ISBN 978-80-7106-982-9.
- HEBAL-JEZIERSKA, M. (2008). Výzkum morfologické variantnosti prostřednictvím ČNK. In Štícha, F. – Fried, M. (eds.). *Grammar & corpora = Gramatika a korpus 2007*. Praha: Academia, 2008, s. 345–351. ISBN 978-80-200-1634-8.
- HEEMSKERK, J. S. (1993). A Probabilistic Context-free Grammar for Disambiguation in Morphological Parsing. In *European Chapter Meeting of the ACL: Proceedings of the sixth conference on European chapter of the Association for Computational Linguistics*. Utrecht,

- 1993, s. 183–192. ISBN 90-5434-014-2. Dostupné na:  
<http://acl.ldc.upenn.edu/E/E93/E93-1023.pdf> [cit. 9. 8. 2010].
- HJELMSLEV, L. (1971). *Jazyk*. Praha: Academia, 1971.
- HJELMSLEV, L. (1972). *O základech teorie jazyka*. Praha: Academia, 1972.
- HLADKÁ, B. – KRÁLÍK, J. (2006). Proměna Českého akademického korpusu. *Slovo a slovesnost*, 2006, roč. 67, č. 3, s. 179–94. ISSN 0037-7031.
- HLADKÁ, Z. (2007). Lexikografie. In Pleskalová, J., et al. (eds.). *Kapitoly z dějin české jazykovědné bohemistiky*. Praha: Academia, 2007, s. 164–198. ISBN 978-80-200-1523-5.
- HLAVÁČKOVÁ, D. – OSOLSOBĚ, K. (2008). Morfologické značkování mluvených korpusů, zkušenosti a otevřené otázky. In Kopřivová, M. – Waclawíčová, M. (eds.). *Čeština v mluveném korpusu*. Praha: Nakladatelství Lidové noviny, 2008, s. 105–114. ISBN 978-80-7106-982-9.
- HLAVÁČOVÁ, J. (2001). Morphological Guesser of Czech Words. In Matoušek, V. – Mautner, P. – Mouček, R. – Taušer, K. (eds.). *Text, Speech and Dialogue: 4th International Conference, TSD 2001, Železná Ruda, Czech Republic, September 11, 2001: Proceedings*. Berlin; Heidelberg; New York: Springer-Verlag, 2001, s. 70–75. ISBN 3-540-42557-8.
- HLAVÁČOVÁ, J. (2009a). *Formalizace systému české morfologie s ohledem na automatické zpracování českých textů*. Praha, 2009. Disertační práce v ÚFAL MFF UK.
- HLAVÁČOVÁ, J. (2009b). Stupňování sloves. In Hlaváčková, D. – Horák, A. – Osolsobě, K. – Rychlý, P. (eds.). *After Half a Century of Slavonic Natural Language Processing*. Brno: MU, 2009, s. 85–90. ISBN 978-80-7399-815-8.
- HLAVÁČOVÁ, J. – HRUŠECKÝ, M. (2008). Affisix: Tool for Prefix Recognition. In Sojka, P. – Horák, A. – Kopeček, I. – Pala, K. (eds.). *Text, Speech and Dialogue, 11th International Conference, TSD 2008, Brno, Czech Republic, September 8–12, 2008, Proceedings*. Berlin; Heidelberg: Springer, 2008, pp. 85–92. ISBN 978-3-540-87390-7.
- HLAVÁČOVÁ, J. – KLÍMOVÁ, J. (2004). Derivational Relations in Flectional Languages – Czech Case. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*. Lisbon, 2004, s. 1239–1242. ISBN 2-9517408-1-6.
- HLAVÁČOVÁ, J. – LOPATKOVÁ, M. (2008). Variants and Homographs: Eternal Problem of Dictionary Makers. In *Proceedings of the 11th International Conference, TSD 2008*. Berlin: Springer, 2008, s. 93–100. ISBN 978-3-540-87390-7. Dostupné na:  
<http://ufal.mff.cuni.cz/~lopatkova/literatura/08-tds-variants.pdf> [cit. 9. 8. 2010].
- HNÁTKOVÁ, M. – JELÍNEK, T. – PETKEVIČ, V. (2009). *Klasifikace a typologie chyb ve vstupních textech a koncepcie značkování chybných textů*. Prezentace k přednášce, 44 slidů, 2009.
- HOCKETT, CH. F. (1967). *A course in modern linguistics*. 12th printing. New York: The Macmillan company, 1967.
- HOLUBOVÁ, V. (2006). Ke studiu variantnosti typu mísen(ý), míšen(ý) aneb Míšení/míšení na bázi korpusu. In Štícha, F. (ed.). *Možnosti a meze české gramatiky*. Praha: Academia, 2006, s. 99–111. ISBN 80-200-1463-2.
- HORECKÝ, J. (1964). *Morfemická štruktúra slovenčiny*. Bratislava: Vydavateľstvo Slovenskej akadémie vied, 1964.
- HORECKÝ, J. (1969). Náčrt fonematickej štruktúry morfém. *Slovenská reč*, 1969, roč. 34, č. 3, s. 137–146. ISSN 0037-6981.
- HORECKÝ, J. (1995). Obsah, forma a funkcia morfémy. *Jazykovedný časopis*, 1995, roč. 46, č. 1, s. 17–22. ISSN 0021-5597.

- HRJAZNUCHINA, T. O., et al. (1979). *Morfemna struktura slova*. Kyjiv: Naukova dumka, 1979.
- CHROMÝ, J. – LEHEČKOVÁ, E. (eds.). (2007). *Rozhovory s českými lingvisty*. Praha: Dauphin, 2007. ISBN 978-80-7272-107-8.
- IVANOVÁ, M. (2006). Afixy – frekvenčná, obsahovo-formálna a funkčná charakteristika. In Sokolová, M. – Ivanová, M. – Ološtiak, M. (eds.). *Morfematický výskum slovenčiny: Možnosti jeho štatistického elektronického a didaktického spracovanie*. Prešov: Filozofická fakulta Prešovskej univerzity v Prešove, 2006, s. 111–144. ISBN 80-8068-544-4.
- IVANOVÁ, M. – GENČI, J. (2006). Morfematicko-derivačné štruktúry lexém. In Sokolová, M. – Ivanová, M. – Ološtiak, M. (eds.). *Morfematický výskum slovenčiny: Možnosti jeho štatistického elektronického a didaktického spracovanie*. Prešov: Filozofická fakulta Prešovskej univerzity v Prešove, 2006, s. 102–110. ISBN 80-8068-544-4.
- JADACKA, H., et al. (2001–2004). *Słownik gniazd słowotwórczych współczesnego języka ogólnopolskiego*. Kraków: Universitas; Towarzystwo Wydawnicze Historia Jagellonica, 2001–2004. Tom 1: Gniazda odprzymiotnikowe. 2001. s. 1–256. Tom 2: Gniazda odrzeczownikowe. s. 264–1410. ISBN 83-7052-591-1. Tom 3: Gniazda odczasownikowe, cz. 1. A–O. 565 s. ISBN 83-88737-02-3. cz. 2. P–Ż. 593 s. ISBN 83-88737-47-3. Tom 4: Gniazda motywowane przez liczebniki, przysłówki, zaimki, przyimki, modulanty, onomatopeje, wykrzykniki. 251 s. ISBN 83-88737-07-4.
- JONGTAVEESATAPORN, M. – WUTIWIWATCHAI, CH. – FURUI, S. (2006). Dictionary Generation Using: Multiple Segmentation Approaches for Thai LVCSR. In *Meeting: Acoustical Society of Japan*. Tokyo, 2006, s. 85–86. ISSN 1880-7658. Dostupné na: <[http://www.furui.cs.titech.ac.jp/publication/2006/0209\\_2-1-8.pdf](http://www.furui.cs.titech.ac.jp/publication/2006/0209_2-1-8.pdf)> [cit. 9. 8. 2010].
- KARPILOVS'KA, J. A. (2002). *Korenovyj gnizdovyj slovnyk ukrajinskoji movi: Hnizda sliv z veršynamy – omohrafičnymy korenjamy*. Kyjiv: Ukrajinska enyklopedija, 2002.
- KESHAHA, S. – PITLER, E. (2005). A simpler, Intuitive Approach to Morpheme Induction. In *Proceedings of 2nd Pascal Challenges Workshop*. Venice (Italy), 2005. Dostupné na: <[http://www.cis.hut.fi/morphochallenge2005/P05\\_KeshavaPitler.pdf](http://www.cis.hut.fi/morphochallenge2005/P05_KeshavaPitler.pdf)> [cit. 9. 8. 2010].
- KLÉGR, A. (2007). *Tezaurus jazyka českého*. Praha: Nakladatelství Lidové noviny, 2007. ISBN 978-80-7106-920-1.
- KLÍMOVÁ, J. (1999). Počítačové zpracování českého odvozování slov. In Nábělková M. – Králik L. (eds.). *Varia VIII*. Bratislava: Slovenská jazykovedná spoločnosť pri SAV, 1999, s. 165–172.
- KLÍMOVÁ, J. (2001). *Computational processing of selected types of czech word-formation*. Summary of Doctoral Thesis, MFF UK, 2001. 18 s.
- KLÍMOVÁ, J. (2010). Český slovtvorný systém 21. století v databázích. In Čmejrková, S. – Hoffmannová, J. – Havlová, E. (eds.). *Užívání a prožívání jazyka: K 90. narozeninám Františka Daneše*. Praha: Karolinum, 2010, s. 147–151. ISBN 978-80-246-1756-5.
- KOLOVRATNÍK, D. – PŘIKRYL, L. (2008). *Programátorská dokumentace k projektu Morfo*. Praha, 2008. Dostupné na: <[http://ufal.mff.cuni.cz/morfo/prog\\_print.pdf](http://ufal.mff.cuni.cz/morfo/prog_print.pdf)> [cit. 9. 8. 2010].
- KOMÁREK, M. (2006a). *Príspevky k českej morfológii*. 2. vyd. Praha: Periplum, 2006. ISBN 80-86624-27-7.
- KOMÁREK, M. (2006b). Formační pravidla české konjugace. In Komárek, M. *Príspevky k českej morfológii*. 2. vyd. Praha: Periplum, 2006, s. 118–154. ISBN 80-86624-27-7.

- KOMÁREK, M. (2006c). K základním pojmům formální morfologie a morfonologie. In Komárek, M. *Příspěvky k české morfologii*. 2. vyd. Praha: Periplum, 2006, s. 95–117. ISBN 80-86624-27-7.
- KOMÁREK, M. (2006d). Ke dvěma koncepcím stavby jednoduchých slovesných tvarů v češtině. In Komárek, M. *Příspěvky k české morfologii*. 2. vyd. Praha: Periplum, 2006, s. 165–171. ISBN 80-86624-27-7.
- KOMÁREK, M. (2006e). Prefixace a slovesný vid: k prefixům prostě vidovým a subsumpci. In Komárek, M. *Příspěvky k české morfologii*. 2. vyd. Praha: Periplum, 2006, s. 183–195. ISBN 80-86624-27-7.
- KOMÁREK, M. (2006f). Prefixální slovesa ve vývoji českého slovesného vidu. In Komárek, M. *Příspěvky k české morfologii*. 2. vyd. Praha: Periplum, 2006, s. 196–203. ISBN 80-86624-27-7.
- KOMÁREK, M. (2006g). Stavba tvarů adjektivní a pronominální flexe v spisovné češtině. In Komárek, M. *Příspěvky k české morfologii*. 2. vyd. Praha: Periplum, 2006, s. 155–164. ISBN 80-86624-27-7.
- KOMÁREK, M. (2006h). Systém autosémantických slovních druhů v spisovné češtině. In Komárek, M. *Příspěvky k české morfologii*. 2. vyd. Praha: Periplum, 2006, s. 13–94. ISBN 80-86624-27-7.
- KOMÁREK, M. (2006i). Změny kmenových konektémů a slovesných typů ve vývoji češtiny. In Komárek, M. *Příspěvky k české morfologii*. 2. vyd. Praha: Periplum, 2006, s. 172–182. ISBN 80-86624-27-7.
- KONEČNÁ, D. – HRONEK, J. (1962). Morfologická analýza podle posledního písmena. *AUC-Phil., SlavPrag*, 1962, č./roč. IV, s. 259–266. ISSN 0323-0813.
- KOPECKÝ, L. V. – FILIPEC, J. – LEŠKA, O., et al. (1976). *Česko-ruský slovník: Díl 2, P–Ž*. Praha: SPN, 1976.
- KOŘENSKÝ, J. (1972). *Komplexní popis výrazové struktury pádu substantiva v češtině*. Praha: Academia, 1972.
- KOUPIL, O. (2007). *Grammatykáři*. Praha: Karolinum, 2007. ISBN 978-80-246-1364-2.
- KRÁLÍK, Ľ. (2006). Synchronia vs. diachronia: niekoľko etymologických poznámok k Slovníku koreňových morfém slovenčiny. In Sokolová, M. – Ivanová, M. – Ološtiak, M. (eds.). *Morfematický výskum slovenčiny: Možnosti jeho štatistického elektronického a didaktického spracovanie*. Prešov: Filozofická fakulta Prešovskej univerzity v Prešove, 2006, s. 284–286. ISBN 80-8068-544-4.
- KRÁLÍK, J. (2009). K formování Českého akademického korpusu. In Hlaváčková, D. – Horák, A. – Osolsobě, K. – Rychlý, P. (eds.). *After Half a Century of Slavonic Natural Language Processing*. Brno: MU, 2009, s. 147–155. ISBN 978-80-7399-815-8.
- KUCHAŘ, J. (1963). Základní rysy struktur pojmenování. *Slovo a slovesnost*, 1963, roč. 24, s. 105–114. ISSN 0037-7031.
- KUČERA, K. (2007). Hyperlemma: A concept Emerging from Lemmatizing Diachronic Corpora In Levinská, J. – Garabík, R. (eds.). [SLOVKO 2007]: Computer Treatment of Slavic and East European Languages. Fourth International Seminar. Bratislava: Tribun, 2007, s. 121–125.
- KURIMO, M. – CREUTZ, M. – VARJOKALLIO, M. (2007). Unsupervised Morpheme Analysis Evaluation by a Comparison to a Linguistic Gold Standard – Morpho Challenge 2007. In *Workshop of the Cross-Language Evaluation Forum, CLEF 2008*. Helsinki, 2007. Dostupné na: <<http://www.cis.hut.fi/mikkok/kurimo1-paperCLEF2008.pdf>> [cit. 9. 8. 2010].



- KUUSIK, E. (1996). Learning Morphology: Algorithms for the Identification of the Stem Changes. In *International Conference On Computational Linguistics: Proceedings of the 16th conference on Computational linguistics*. Vol. 2. Copenhagen (Denmark), 1996, s. 1102–1105. Dostupné na: <<http://www.aclweb.org/anthology/C/C96/C96-2198.pdf>> [cit. 9. 8. 2010].
- KUZNĚCOVA, A. I. – JEFREMOVA, T. F. (1986). *Slovar' morfem russkogo jazyka: Okolo 52 000 slov*. Moskva: Rus. jazyk, 1986.
- LEBEDA, J. (2009). Nedostizitelný Josef Jungmann. *Lidové noviny*, 31. 10. 2010, Orientace, s. 24. ISSN 1213-1385.
- LEE, G. G. – CHA, J. – LEE, J. (2002). Syllable pattern-based unknown morpheme segmentation and estimation for hybrid part-of-speech tagging of Korean. *Computational Linguistics*, 2002, vol. 28, issue 1, s. 53–70. ISSN 0891-2017. Dostupné na: <<http://nlp.postech.ac.kr/Course/CS730b/2001/postag.ps>> [cit. 9. 8. 2010].
- LEE, Y.-S. – PAPINENI, K. – ROUKOS, S. – EMAM, O. – HASSAN, H. (2003). Language Model Based Arabic Word Segmentation. In *Annual Meeting of the ACL: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*. Vol. 1. Sapporo (Japan), 2003, s. 399–406. Dostupné na: <<http://acl.ldc.upenn.edu/acl2003/main/pdfs/LeePapineni.pdf>> [cit. 9. 8. 2010].
- LEŠKA, O. (1993). K Novákovým „osudům české lingvistiky“. *Slovo a slovesnost*, 1993, roč. 54, č. 2, s. 132–134. ISSN 0037-7031.
- LEZIUS, W. – RAPP, R. – WETTLER, M. (1998). A Freely Available Morphological Analyzer, Disambiguator and Context Sensitive Lemmatizer for German. In *Annual Meeting of the ACL: Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*. Vol. 2. Montreal, 1998, s. 743–748. Dostupné na: <<http://www.aclweb.org/anthology/P/P98/P98-2123.pdf>> [cit. 9. 8. 2010].
- MANSOUR, S. (2008). *Combining Character and Morpheme Based Models for Part-Of-Speech Tagging of Semitic Languages*. M.Sc. Thesis. Haifa, 2008. Dostupné na: <<http://www.cs.technion.ac.il/events/2007/237/>> [cit. 9. 8. 2010].
- MATHÉ, I. – ŠPAČEK, L. (2010). *Etiketa*. 2. vyd. Praha: BB/art. ISBN 978-80-7381-719-0.
- MARCUS, S. (1969). *Algebraické modely v lingvistice*. Praha: Academia, 1969.
- MARVAN, J. (2006). *Cesty ke spisovné češtině (800–1800)*. Ústí nad Labem: PedF UJEP, 2006. ISBN 80-7044-804-0.
- MARVAN, J. (2008). *České stupňování – slovanská podoba evropské univerzálie*. Praha: Karolinum, 2008. ISBN 978-80-246-1301-7.
- MACHAČ, P. – SKARNITZL, R. (2009). *Fonetická segmentace hlásek*. Praha: Epocha, 2009. ISBN 978-80-7425-031-6.
- MARVAN, J. (2008). *České stupňování – slovanská podoba evropské univerzálie*. Praha: Karolinum, 2008. ISBN 978-80-246-1301-7.
- MANNING, CH. D. – SCHÜTZE, H. (1999). *Foundations of statistical natural language processing*. Cambridge, Mass.: MIT Press, 1999. ISBN 0-262-13360-1.
- MEL'ČUK, I. (2006). *Aspects of the theory of morphology*. Berlin; New York: Mouton de Gruyter, 2006. ISBN 978-3-11-017711-4.
- MENGEL, A. (1999). A phonetic morpheme lexicon for german. In *Proceedings of the International Conference on Phonetic Sciences ICPH'99*. San Francisco, 1999. Dostupné na: <<http://www.andreasmengel.de/pubs/icphs-sf-99-morf.pdf>> [cit. 9. 8. 2010].

- [MČ1] DOKULIL, M., et al. (reds.). (1986). *Mluvnice češtiny 1: Fonetika, fonologie, morfonologie a morfémiologie, tvoření slov*. Praha: Academia, 1986.
- [MČ2] KOMÁREK, M., et al. (reds.). (1986). *Mluvnice češtiny 2: Tvarosloví*. Praha: Academia, 1986.
- MIKHEEV, A. (1997). Automatic Rule Induction for Unknown-Word Guessing. *Computational Linguistics*, 1997, vol. 23, issue 3, s. 405–423. ISSN 0891-2017. Dostupné na: <<http://acl.ldc.upenn.edu/J/J97/J97-3003.pdf>> [cit. 9. 8. 2010].
- MITTER, P. (2003a). Hybridní kompozita v současné češtině z hlediska slohových příznaků a expresivity. *Bohemistika*, 2003, č. 4, s. 321–327. ISSN 1642-9893.
- MITTER, P. (2003b). *Složená hybridní substantiva s prvním komponentem cizího původu v současné češtině*. Ústí nad Labem: Univerzita J. E. Purkyně, 2003. ISBN 80-7044-554-8.
- MITTER, P. (2004a). Cizí první komponenty složených slov z pohledu sémantiky. *Bohemistika*, 2004, č. 3, s. 196–202. ISSN 1642-9893.
- MITTER, P. (2004b). Hybridní kompozita v českých mluvnicích 2. poloviny 20. století z hlediska spisovnosti. In Minářová, E. – Ondrášková, K. (eds.). *Spisovnost a nespisovnost: Zdroje, proměny a perspektivy*. Brno: PedF MU, 2004, s. 181–187.
- MITTER, P. (2005a). Kompozitní komponenty internacionálního charakteru v současné české slovní zásobě a jejich domácí ekvivalenty. In Uličný, O. (ed.). *Euroletteraria & Eurolingua 2005*, Opera Academiae Paedagogicae Liberecensis, Vol. III. Liberec, 2005, s. 273–284. ISBN 80-7372-040-X.
- MITTER, P. (2005b). Vymezení některých prefixoidů a radixoidů v české lingvistické literatuře. In Uličný, O. (ed.). *Eurolingua 2004*, Opera Academiae Paedagogicae Liberecensis, Vol. III. Liberec, 2005, s. 57–62. ISBN 80-7083-958-9.
- MITTER, P. (2006). *Kompozice v kontextu současné češtiny*. Ústí nad Labem: Univerzita Jana Evangelisty Purkyně v Ústí nad Labem, 2006. ISBN 80-7044-811-3.
- MITTER, P. (2007). K jednomu typu hybridních složenin ve druhém svazku neologického slovníku. *Bohemistika*, 2007, č. 1, s. 31–42. ISSN 1642-9893.
- MITTER, P. (2008). *Hybridní složeniny s prvním komponentem domácího původu v současné češtině*. Ústí nad Labem: Univerzita J. E. Purkyně, 2008. ISBN 978-80-7417-080-8.
- MONSON, CH. – CARBONELL, J. – LAVIE, A. – LEVIN, L. (2007). ParaMor: Finding Paradigms across Morphology. In *Advances in Multilingual and Multimodal Information Retrieval*. Berlin: Springer, 2008, s. 900–907. ISBN 978-3-540-85759-4. Dostupné na: <[http://www.cs.cmu.edu/~jgc/publication/ParaMor\\_Finding\\_CLEF\\_2007.pdf](http://www.cs.cmu.edu/~jgc/publication/ParaMor_Finding_CLEF_2007.pdf)> [cit. 9. 8. 2010].
- MONSON, CH. (2004). A Framework for Unsupervised Natural Language Morphology Induction. In *Annual Meeting of the ACL: Proceedings of the ACL 2004 workshop on Student research*. Barcelona, 2004. Dostupné na: <<http://acl.ldc.upenn.edu/P/P04/P04-2012.pdf>> [cit. 9. 8. 2010].
- MORRIS, CH. (1946). *Signs, language and behavior*. New York: George Braziller, 1946.
- MORRIS, CH. W. (1997). Základy teorie znaku. In Palek, B. (ed.). *Sémiotika*. 2., přepr. vyd. Praha: Karolinum, 1997, s. 197–256. ISBN 80-7184-356-3.
- MRAVINACOVÁ, J. (2003). Anomální kombinace internacionálních morfémů s prvky českého lexika. In Rangelova, A. – Tichá, Z. *Internacionalizmy v nové slovní zásobě*. Praha: ÚJČ AV ČR, 2003, s. 126–132. ISBN 80-86496-11-2.
- NAKOV, P. – ANGELOVA, G. – HAHN, W. von. (2001). Automatic Recognition and Morphological Classification of Unknown German Nouns. In *Universität Hamburg*,

- Fachbereich Informatik: Bericht 243*. 2002, 53 s. Dostupné na: <<http://nats-www.informatik.uni-hamburg.de/~vhahn/Downloads/Report243.pdf>> [cit. 9. 8. 2010].
- NEBESKÁ, I. (1992). *Úvod do psycholingvistiky*. Jinočany: H&H, 1992. ISBN 80-85467-75-5.
- [NASCS] KRAUS, J. – PETRÁČKOVÁ, V., et al. (2005). *Nový akademický slovník cizích slov*. Praha: Academia, 2005. ISBN 80-200-1351-2.
- OFLAZER, K. – NIRENBURG, S. – MCSHANE, M. (2001). Bootstrapping Morphological Analyzers by Combining Human Elicitation and Machine Learning. *Computational Linguistics*, 2001, vol. 27, issue 1, s. 59–85. ISBN 0891-2017. Dostupné na: <<http://acl.ldc.upenn.edu/J/J01/J01-1003.pdf>> [cit. 9. 8. 2010].
- OGDEN, C. K. – RICHARDS, L. A. (1997). Myšlenky, slova a věci. In Palek, B. (ed.). *Sémiotika*. 2., přepr. vyd. Praha: Karolinum, 1997, s. 173–195. ISBN 80-7184-356-3.
- OLIVA, K., jr. – HNÁTKOVÁ, M. – PETKEVIČ, V. – KVĚTOŇ, P. (2000). The Linguistic Basis of a Rule-Based Tagger of Czech. In *Text, Speech and Dialogue*. Berlin: Springer, 2000, s. 205–232. ISBN 978-3-540-41042-3.
- OLOŠTIAK, M. (2006). Štatistické spracovanie propriálných radixov. In Sokolová, M. – Ivanová, M. – Ološtiak, M. (eds.). *Morfematický výskum slovenčiny: Možnosti jeho štatistického elektronického a didaktického spracovanie*. Prešov: Filozofická fakulta Prešovskej univerzity v Prešove, 2006, s. 57–78. ISBN 80-8068-544-4.
- ONG, W. J. (2006). *Technologizace slova: Mluvená a psaná řeč*. Praha: Karolinum, 2006. ISBN 80-246-1124-4.
- OSOLSOBĚ, K. (1996). *Algoritmický popis české formální morfologie a strojový slovník češtiny*. Brno, 1996. Disertační práce na FF MU.
- OSOLSOBĚ, K. (2006). Korpus soukromé korespondence (KSK) z hlediska morfologického značkování. In *SPFFMU*. Brno: MU, 2006, s. 185–199. ISBN 80-210-4014-9.
- OSOLSOBĚ, K. (2007a). Korpus jako zdroj dat pro opravy chyb automatické morfologické analýzy. In *Grammar & Corpora, 2nd International Conference, Abstracts*. Praha: ÚJČ AV ČR, 2007, s. 59–61. ISBN 80-86496-35-X.
- OSOLSOBĚ, K. (2007b). Popis gramatických významů (hodnot) jednoduchých slovesných tvarů v anotacích českých (slovenských) korpusů. *Linguistica Brunensia*, A 55, č. 1, Brno: MU, 2007, s. 201–218. ISSN 0231-7567.
- OSOLSOBĚ, K. (2008a). Značkování a status některých gramatických kategorií v ČNK (syntetické futurum, stupňování adjektiv, neurčité číslovky a příslovce míry). In Štícha, F. – Fried, M. (eds.). *Grammar & corpora = Gramatika a korpus 2007*. Praha: Academia, 2008, s. 407–416. ISBN 978-80-200-1634-8.
- OSOLSOBĚ, K. ([2008b]). *Automatický nástroj Deriv*. Dostupné na: <[http://is.muni.cz/el/1421/jaro2008/CJBB84/Automaticky\\_nastroj\\_Deriv.doc](http://is.muni.cz/el/1421/jaro2008/CJBB84/Automaticky_nastroj_Deriv.doc)> [cit. 9. 8. 2010].
- PACHUNKE, T. – MERTINEIT, O. – WOTHE, K. – SCHMIDT, R. (1992). Broad Coverage Automatic Morphological Segmentation of German Words. In *International Conference On Computational Linguistics: Proceedings of the 14th conference on Computational linguistics*. Vol. 4. Nantes (France), 1992, s. 1218–1222. Dostupné na: <<http://acl.ldc.upenn.edu/C/C92/C92-4195.pdf>> [cit. 9. 8. 2010].
- PALA, K. – HLAVÁČKOVÁ, D. (2007). Computer Processing Derivational Relations in Czech. In Levinská, J. – Garabík, R. (eds.). [SLOVKO 2007]: Computer Treatment of Slavic and East European Languages. Fourth International Seminar. Bratislava: Tribun, 2007, s. 198–208.

- PALA, K. – SEDLÁČEK, R. – VEBER, M. (2004). Vztah mezi tvarotvornými a slovotvornými vzory v češtině. In Hladká, Z. – Karlík, P. (eds.). *Čeština – univerzália a specifiká* 5. Praha: Nakladatelství Lidové noviny, 2004, s. 151–162. ISBN 80-7106-659-1.
- PALEK, B. (1989). *Základy obecné jazykovědy*. Praha: Státní pedagogické nakladatelství, 1989. ISBN 80-04-22937-9.
- PETKEVIČ, V. (2006). Reliable morphological disambiguation of Czech: a rule-based approach is necessary. In Šimková, M. (ed.). *Insight into the Slovak and Czech Corpus Linguistics*. Bratislava: Veda, pp. 26–44. ISBN 80-224-0880-8.
- PETR, J. (1988). F. L. Čelakovský a slovanská jazykověda. In Petr, J. – Urban, Z. (eds.). *Slavistický odkaz F. Čelakovského. Práce z dějin slavistiky*, XI. Praha: UK, 1988, s. s. 5–62.
- PIASECKI, M. – RADZISZEWSKI, A. (2007). Polish Morphological Guesser Based on a Statistical A Tergo Index. In *Proceedings of the International Multiconference on Computer Science and Information Technology*. 2007, s. 247–256. ISSN 1896-7094. Dostupné na: <<http://www.plwordnet.pwr.wroc.pl/main/content/files/publications/aaia07-mp-ar150.pdf>> [cit. 9. 8. 2010].
- PINKER, S. (2008). *Slova a pravidla: Složky jazyka*. Praha: Academia, 2008. ISBN 978-80-200-1641-6.
- PLESKALOVÁ, J., et al. (eds.). (2007). *Kapitoly z dějin české jazykové bohemistiky*. Praha: Academia, 2007, s.273– 277. ISBN 978-80-200-1523-5.
- POLIKARPOV, A. A. (2000). Chronological Morphemic and Word-Formational Dictionary of Russian: Some System Regularities for Morphemic Structures and Units. In Zybatow, G. – Junghanns, U. – Mehlhorn, G. – Szucsich, L. (eds.). *3. Europäische Konferenz „Formale Beschreibung slavischer Sprachen“*, Leipzig, 1999. Leipzig, 2000, s. 201–212.
- POMIKÁLEK, J. – RYCHLÝ, P. – KILGARRIFF, A. (2009). Scaling to Billion-plus Word Corpora. In Gelbukh, A. (ed.) *Advances in Computational Linguistics. Research in Computing Science*. 41, 2009, s. 3–14. ISSN 1870-4069. Dostupné na: <<http://site.cicling.org/2009/RCS-41/003-014.pdf>> [cit. 9. 8. 2010].
- PSUTKA, J., et al. (2006). *Mluvíme s počítačem česky*. Praha: Academia, 2006. ISBN 80-200-1309-1.
- RAPPOPORT, A. – LEVENT-LEVI, T. (2006). Induction of Cross-Language Affix and Letter Sequence Correspondence. In *ACL Workshops: Proceedings of the International Workshop on Cross-Language Knowledge Induction*. Trento (Italy), 2006, s. 17–24. Dostupné na: <<http://www.mt-archive.info/EACL-2006-Rappoport.pdf>> [cit. 9. 8. 2010].
- REHMAN, K. U. – HUSSAIN, I. (2005). Unsupervised Morphemes Segmentation. In *Proceedings Morpho Challenge 2005*. Dostupné na: <[http://www.cis.hut.fi/morphochallenge2005/P10\\_RehmanHussain.pdf](http://www.cis.hut.fi/morphochallenge2005/P10_RehmanHussain.pdf)> [cit. 9. 8. 2010].
- REICHEL, U. D. – WEILHAMMER, K. (2004). Automated Morphological Segmentation and Evaluation. In *Proceedings of LREC 2004*. Lisbon, 2004, s. 503–506. Dostupné na: <[http://www.phonetik.uni-muenchen.de/~reichelu/reichel\\_weilhammer\\_paper.ps](http://www.phonetik.uni-muenchen.de/~reichelu/reichel_weilhammer_paper.ps)> [cit. 9. 8. 2010].
- REJZEK, J. (2001). *Český etymologický slovník*. Voznice: Leda, 2001. ISBN 80-85927-85-3.
- REJZEK, J. (2009). *Lidová etymologie v češtině*. Praha: Karolinum, 2009. ISBN 978-80-246-1649-0.
- RIESA, J. – YAROWSKY, D. (2006). Minimally Supervised Morphological Segmentation with Applications to Machine Translation. In *Proceedings of the 7th Conference of the*



- Association for Machine Translation in the Americas (AMTA). 2006. Dostupné na: <[http://www.isi.edu/~riesa/pubs/morph\\_amta06.pdf](http://www.isi.edu/~riesa/pubs/morph_amta06.pdf)> [cit. 9. 8. 2010].
- [RMSČ] SLAVÍČKOVÁ, E. (1975). *Retrográdní morfemický slovník češtiny: s připojenými inventárními slovníky českých morfémů kořenových, prefixálních a sufixálních*. Praha: Academia, 1975.
- RODOVICH, N. (1971). *Glossario morfematico dello slavo ecclesiastico antico*. Napoli: Edizioni Cymba, 1971.
- ROMPORTL, S. (1970). *Struktura gramatické složky slovesných tvarů určitých v češtině*. Praha. Academia, 1970.
- RONZHIN, A. L. – KARPOV, A. A. (2004). Implementation of morphemic analysis for russian speech recognition. In *SPECOM 2004*. St. Petersburg, 2004, s. 291–296. Dostupné na: <[http://www.isca-speech.org/archive/specom\\_04/spc4\\_291.pdf](http://www.isca-speech.org/archive/specom_04/spc4_291.pdf)> [cit. 9. 8. 2010].
- RYBA, B. (1949). O Komenského neznámém spisku „Kořenové jazyka českého“. *Listy filologické*, 1949, roč. 73, s. 171–178. ISSN 0024-4457.
- SABOL, J. (1973). O frekvencii morfém v slovenčine. *Slovenská reč*, 1973, roč. 38, č. 3–4, s. 138–143. ISSN 0037-6981.
- SABOL, J. (1979). Konsonanticko-vokalické zloženie morfém v spisovnej slovenčine. *Jazykovedný časopis*, 1979, roč. 20, č. 2, s. 120–133. ISSN 0021-5597.
- SABOL, J. – BÓNOVÁ, I. (2003). Štruktúra slabiky a morfémy v slovenčine, češtině a polštině: Metodologická východiská výskumu. In *XIII. Medzinárodný zjazd slavistov v Lubľane. Príspevky slovenských slavistov*. Bratislava, 2003, s. 101–109. ISBN 80-968971-2-8.
- SABOL, J. – BÓNOVÁ, I. (2004). K problematike vzťahu slabiky a morfémy v slovenčine. In Tatár, J. (ed.). *Studia Slovaca: Zborník vedeckých prác členov Katedry slovenského jazyka a literatúry*. Banská Bystrica: PedF UMB, 2004, s. 199–204. ISBN 80-8055-948-1.
- SAUSSURE, F. de. (1996). *Kurs obecné lingvistiky*. 2. vyd. Praha: Academia, 1996. ISBN 80-200-0560-9.
- SEDLÁČEK, R. (1999). *Morfologický analyzátor češtiny*. Brno, 1999. Diplomová práce na Fakultě informatiky MU.
- SEDLÁČEK, R. (2004). *Morphemic analyser for Czech*. Brno, 2004. PhD thesis at Faculty of informatics MU.
- SGALL, P. (1960). Soustava pádových koncovek v češtině. *AUC-Phil., SlavPrag*, 1960, roč./č. II, s. 65–83. ISSN 0323-0813
- SGALL, P. (1967). *Generativní popis jazyka a česká deklinace*. Praha: Academia, 1967.
- SCHONE, P. – JURAFSKY, D. (2000). Knowledge-Free Induction of Morphology Using Latent Semantic Analysis. In *Proceedings of CoNLL-2000 and LLL-2000*. Dostupné na: <<http://www.aclweb.org/anthology-new/W/W00/W00-0712.pdf>> [cit. 9. 8. 2010].
- SCHWARZOVÁ, M. (2009). *Úvod do kognitivní lingvistiky*. Praha: Dauphin, 2009. ISBN 978-80-7272-155-9.
- SKALIČKA, V. (1935). Asymetrický dualismus jazykových jednotek. *Naše řeč*, 1935, roč. 19, č. 6–7, s. 296–303. ISSN 0027-8203.
- SKALIČKA, V. (2004a). Komplexnost jazykových jednotek. In Čermák, F. – Čermák, J. – Čermák, P. – Poeta, C. (eds.). *Vladimír Skalička: Souborné dílo II*. Praha: Karolinum, 2004, s. 612–620. ISBN 80-246-0734-4.

- SKALIČKA, V. (2004b). O pojem morfému. In Čermák, F. – Čermák, J. – Čermák, P. – Poeta, C. (eds.). *Vladimír Skalička: Souborné dílo I*. Praha: Karolinum, 2004, s. 179–188. ISBN 80-246-0549-X.
- SKALIČKA, V. (2004c). O roli vnitřní flexe v jazyce. In Čermák, F. – Čermák, J. – Čermák, P. – Poeta, C. (eds.). *Vladimír Skalička: Souborné dílo I*. Praha: Karolinum, 2004, s. 376–384. ISBN 80-246-0549-X.
- SKOUMALOVÁ, Z. (1976). *Status tzv. kmenotvorných přípon slovesných v slovanských jazycích, zvláště v ruštině*. Praha: Kabinet cizích jazyků ČSAV, 1976.
- SKOUMALOVÁ, H. (1997). A Czech Morphological Lexicon. In *Proceedings of the Third Meeting of the ACL: Special Interest Group in Computational Phonology*. Madrid, 1997, s. 41–47. Dostupné na: <<http://www.aclweb.org/anthology/W/W97/W97-1106.pdf>> [cit. 9. 8. 2010].
- SLAVÍČKOVÁ, E. (1962). Metoda morfémové analýzy založená na aplikaci teorie pravděpodobnosti. *Slovo a slovesnost*, 1962, roč. 23, č. 2, s. 94–104. ISSN 0037-7031.
- SLAVÍČKOVÁ, E. (1965). Rozbor a kvantitativní hodnocení českých kořenových morfémů. In *Problémy kybernetiky*. Praha: Nakladatelství Československé akademie věd, 1965, s. 360–371.
- SLAVÍČKOVÁ, E. (1967). Některé problémy morfémové analýzy češtiny. *Slovo a slovesnost*, 1967, roč. 28, č. 1, s. 7–17. ISSN 0037-7031.
- SLAVÍČKOVÁ, E. – VAŠÁK, P. (1972). Statistical examination of czech root-morphemes. *Prague Studies in Mathematical Linguistics*, 1972, [sv.] 3, s. 89–101.
- SMOLÍK, F. (2009). Psycholingvistika a čeština: některá slibná témata. *Naše řeč*, 2009, roč. 92, č. 5, s. 240–251. ISSN 0027-8203.
- SNOVER, M. G. – JAROSZ, G. E. – BRENT, M. R. (2002). Unsupervised Learning of Morphology Using a Novel Directed Search Algorithm: Taking the First Step. In *Morphological and Phonological Learning: Proceedings of the 6th Workshop of the ACL Special Interest Group in Computational Phonology (SIGPHON)*. Philadelphia, 2002, s. 11–20. Dostupné na: <<http://www.aclweb.org/anthology/W/W02/W02-0602.pdf>> [cit. 9. 8. 2010].
- SOJKA, P. (2005). *Competing Patterns in Language Engineering and Computer Typesetting*. Brno, 2005. Dissertation at FI MU. Dostupné na: <<http://www.fi.muni.cz/usr/sojka/papers/sojka-thesis.pdf>> [cit. 9. 8. 2010].
- SOKOLOVÁ, M. (1994). Segmentácia prevzatých slov v slovenčine. *Jazykovedný časopis*, 1994, roč. 45, č. 2, s. 128–149. ISSN 0021-5597.
- SOKOLOVÁ, M. (1999). Teoretické zásady morfematického spracovania slovenčiny. In Sokolová, M. – Moško, G. – Šimon, F. – Benko, V. *Morfematický slovník slovenčiny*. Prešov: Náuka, 1999, s. 9–56. ISBN 80-968202-1-4.
- SOKOLOVÁ, M. (2006). Morfematika vo vyučovaní slovenského jazyka na vysokých školách. In Sokolová, M. – Ivanová, M. – Ološtiak, M. (eds.). *Morfematický výskum slovenčiny: Možnosti jeho štatistického elektronického a didaktického spracovanie*. Prešov: Filozofická fakulta Prešovskej univerzity v Prešove, 2006, s. 177–234. ISBN 80-8068-544-4.
- SOKOLOVÁ, M. (2008). Morfematický výskum slovenčiny medzi XIII. a XIV. slavistickom zjazdom. *Jazykovedný časopis*, 2008, roč. 59, č. 1–2, s. 25–36. ISSN 0021-5597.
- SOKOLOVÁ, M. – GENČI, J. (2006a). Fonematické štruktúry radixov. In Sokolová, M. – Ivanová, M. – Ološtiak, M. (eds.). *Morfematický výskum slovenčiny: Možnosti jeho*

- štatistického elektronického a didaktického spracovanie*. Prešov: Filozofická fakulta Prešovskej univerzity v Prešove, 2006, s. 79–101. ISBN 80-8068-544-4.
- SOKOLOVÁ, M. – GENČI, J. (2006b). Štatistické spracovanie apelatívnych radixov. In Sokolová, M. – Ivanová, M. – Ološtiak, M. (eds.). *Morfematický výskum slovenčiny: Možnosti jeho štatistického elektronického a didaktického spracovanie*. Prešov: Filozofická fakulta Prešovskej univerzity v Prešove, 2006, s. 27–56. ISBN 80-8068-544-4.
- SOKOLOVÁ, M. – IVANOVÁ, M. (2007). The heritage of the linguistic work by Ján Horecký in contemporary morphemic research of the Slovak language. *SKASE Journal of Theoretical Linguistics*, 2007, vol. 4, no. 1, pp. 69–85. ISSN 1336-782X.
- SOKOLOVÁ, M. – IVANOVÁ, M. – DZIAKOVÁ, K. (2003). Zásady spracovania slovníka morfév v slovenčine. *Jazykovedný časopis*, 2003, roč. 54, č. 1–2, s. 67–92. ISSN 0021-5597.
- SOKOLOVÁ, M. – IVANOVÁ, M. – OLOŠTIAK, M. (eds.). (2006). *Morfematický výskum slovenčiny: Možnosti jeho štatistického elektronického a didaktického spracovanie*. Prešov: Filozofická fakulta Prešovskej univerzity v Prešove, 2006. ISBN 80-8068-544-4.
- SOKOLOVÁ, M. – IVANOVÁ, M. – VUŽŇÁKOVÁ, K. (2005). Prínos slovníka koreňových morfév k charakteristike kompozít. *Jazykovedný časopis*, 2005, roč. 56, č. 2, s. 99–126. ISSN 0021-5597.
- SOKOLOVÁ, M. – MOŠKO, G. – ŠIMON, F. – BENKO, V. (1999). *Morfematický slovník slovenčiny*. Prešov: Náuka, 1999. ISBN 80-968202-1-4.
- SOKOLOVÁ, M. – OLOŠTIAK, M. – IVANOVÁ, M. (eds.). (2005). *Slovník koreňových morfév slovenčiny*. Prešov: FF Prešovskej univerzity, 2005. 584 s. ISBN 80-8068-319-0.
- SOKOLOVÁ, M. – ŠIMON, F. – MOŠKO, G. (1996). Zásady morfematickej analýzy v morfematickom slovníku slovenčiny. *Jazykovedný časopis*, 1996, roč. 47, č. 1, s. 40–50. ISSN 0021-5597.
- SOLNCEV, V. M. (1981). *Systém a struktura v jazyce*. Praha: Academia, 1981.
- SOLOV'JEV, S. M. – SNARSKAJA, S. M. (2006). *Morfemnyj i slovoobrazovatelnyj razbor*. Sankt-Peterburg: Norint, 2006. ISBN 5-7711-0169-9.
- SPOUSTA, M. (2005). *Automatické přiřazení tvaroslovných vzorů v češtině*. Praha, 2005. Diplomová práce v Ústavu formální a aplikované lingvistiky MFF UK.
- SPOUSTOVÁ, D. (2007). *Kombinované statisticko-pravidlové metody značkování češtiny*. Praha, 2007. Disertační práce na ÚFAL MFF UK.
- SPOUSTOVÁ, D. – HAJIČ, J. – VOTRUBEC, J. – KRBEC, P. – KVĚTOŇ, P. (2007). The Best of Two Worlds: Cooperation of Statistical and Rule-Based Taggers for Czech. In *ACL 2007 workshop: Proceedings of the Workshop on Balto-Slavonic Natural Language Processing 2007*. 2007, s. 67–74. ISBN 978-1-932432-88-6. Dostupné na: <http://ufal.mff.cuni.cz/czech-tagging/JohankaEtAlBSNLP2007.pdf> [cit. 9. 8. 2010].
- SPOUSTOVÁ, D. – HAJIČ, J. – RAAB, J. – SPOUSTA, M. (2009). Semi-supervised Training for the Averaged Perceptron POS Tagger. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*. Athina (Greece), 2009, s. 763–771. ISBN 978-1-932432-16-9.
- SPOUSTOVÁ, D. – SPOUSTA, M. – PECINA, P. (2010). Building a web corpus of Czech. In *Proceedings of the seventh International conference on Language Resources and Evaluation (LREC 2010)*. Malta: Valletta, 2010, s. 998–1001.
- [SSČ] MEJSTŘÍK, V. (hl. red.), et al. (2003). *Slovník spisovné češtiny pro školu a veřejnost*. 3., opr. vyd. Praha: Academia. 1. vyd. 1978, 2., opr. a dopl. vyd. 1994. ISBN 80-200-1080-7.

- [SSJČ] HAVRÁNEK, B. (hl. red.), et al. (1989). *Slovník spisovného jazyka českého: 1–8*. 2., nezm. vyd. Praha: Academia, 1989. ISBN 978-80-7106-920-1.
- [StatČ] BARTOŇ, T. – CVRČEK, V. – ČERMÁK, F. – JELÍNEK, T. – PETKEVIČ, V. (2009). *Statistiky češtiny*. Praha: Nakladatelství Lidové noviny, 2009. ISBN 978-80-7106-594-4.
- STICH, A. (2001). Dopisování strukturních lingvistů. *Souvislosti*, 2001, roč. 12, č. 3–4, s. 216–222.
- STRAKOVÁ, V. (1973). *Substantivní derivace v ruštině a češtině*. Praha: Academia, 1973.
- STRAKOVÁ, V. (1985). *Morfematická strukturace sémantických obsahů: Nástin derivační typologie*. Praha; Třinec: ÚJČ ČSAV, 1985.
- SUNITHA, K. V. N. – KALYANI, N. (2009). Improving word coverage using unsupervised morphological analyser. *Sādhāna*, vol. 34, part 5, October 2009, pp. 703–715.
- ŠAFAŘÍK, P. J. (1846). O tvoření slov zdvojováním kořene. *Časopis Českého museum*, 1846, č. 3, s. 355–368.
- ŠAFAŘÍK, P. J. (1848). Mluvozpýtný rozbor čísl slova. *Časopis Českého museum*, 1848, č. 3, s. 217–257.
- ŠIMANDL, J. (2010). *Dnešní skloňování substantiv typů kámen, břímě*. Praha: Nakladatelství Lidové noviny, 2010. ISBN 978-80-7422-008-1.
- ŠIMON, F. (2006). Alternácie a varianty koreňových morfém prevzatých slov. In Sokolová, M. – Ivanová, M. – Ološtiak, M. (eds.). *Morfematický výskum slovenčiny: Možnosti jeho štatistického elektronického a didaktického spracovanie*. Prešov: Filozofická fakulta Prešovskej univerzity v Prešove, 2006, s. 165–174. ISBN 80-8068-544-4.
- ŠIŠKA, Z. (2005). *Bázový morfematický slovník češtiny*. 2. vyd. Olomouc: PedF UP, 2005. 227 s. 1. vyd. 1998. ISBN 80-244-0962-3.
- ŠLOSAR, D. (1977). [Slavičková, E.:] Retrográdní morfematický slovník češtiny. *SPFFBU*, A, 1977, roč. 25–26, s. 178–179.
- ŠMERK, P. (2007a). K morfologické desambiguaci češtiny. Brno, 2007. Teze disertační práce. Dostupné na: <[http://is.muni.cz/th/3880/fi\\_r/teze.pdf](http://is.muni.cz/th/3880/fi_r/teze.pdf)> [cit. 9. 8. 2010].
- ŠMERK, P. (2007b). Morphemic Analysis: A dictionary Lookup Instead of Real Analysis. In Sojka, P. – Horák, A. *First Workshop on Recent Advances in Slavonic Natural Language Processing, RASLAN 2007*. Brno: Masaryk University, 2007, s. 77–85. ISBN 978-80-210-4471-5.
- ŠMERK, P. (2008). Towards Czech Morphological Guesser. In Sojka, P. – Horák, A. *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2008*. Brno: Masarykova univerzita, 2008. s. 1–4, 4 s. ISBN 978-80-210-4741-9.
- ŠMERK, P. (2009). Fast morphological analysis of czech. In Sojka, P. – Horák, A. (eds.). *Proceedings of recent advances in slavonic natural language processing, RASLAN 2009*. Brno: MU, 2009, pp. 6–10. ISBN 978-80-7399-815-8.
- ŠMERK, P. (2010). *K počítačové morfologické analýze češtiny*. Disertační práce na KIT, FI MU, 2010.
- ŠRÁMEK, R. (1999). *Úvod do ebecné onomastiky*. Brno: Masarykova univerzita, 1999. ISBN 80-210-2027-X.
- ŠTINDLOVÁ, J. (1961). Stroje na zpracování informací a jejich význam pro jazykovědu. *Naše řeč*, 1961, roč. 44, s. 208–215. ISSN 0027-8203.
- ŠTINDLOVÁ, J. (1968). Dělení slov v češtině pomocí strojů. *Naše řeč*, 1968, roč. 51, s. 23–33. ISSN 0027-8203.



- TĚŠITELOVÁ, M., et al. (1985). *Kvantitativní charakteristiky současné češtiny*. Praha: Academia, 1985.
- TĚŠITELOVÁ, M. (1999). K úloze časopisu Slovo a slovesnost ve vývoji kvantitativní lingvistiky. *Slovo a slovesnost*, 1999, roč. 60, č. 2, s. 81–87.
- [Teze] (1929/1970). Teze předložené prvému sjezdu slovanských filologů v Praze 1929. In Vachek, J. (ed.). *U základů pražské jazykovědné školy*. Praha: Academia, 1970, s. 35–65.
- THURMAIR, G. (1984). Linguistic Problems in Multilingual Morphological Decomposition. In *Annual Meeting of the ACL: Proceedings of the 10th International Conference on Computational Linguistics and 22nd annual meeting on Association for Computational Linguistics*. Stanford, 1984, s. 174–177. Dostupné na: <<http://acl.ldc.upenn.edu/P/P84/P84-1037.pdf>> [cit. 9. 8. 2010].
- TICHONOV, A. H. (1996). *Morfemno-orfografičeskij slovar' russkogo jazyka: Ruskaja morphemika*. Moskva: Škola-Press, 1996. 704 s. ISBN 5-88527-119-4.
- TRNKA, B. (1990a). Jazykové roviny a jejich strategie. In Trnka, B. *Kapitoly z funkční jazykovědy*. Praha: UK, 1990, s. 19–26.
- TRNKA, B. (1990b). Zásady strukturní morfologie. In Trnka, B. *Kapitoly z funkční jazykovědy*. Praha: UK, 1990, s. 143–150.
- TUŠKOVÁ, J. M. (2006). Variantní tvary v deklinaci vzoru růže. In Šticha, F. (ed.). *Možnosti a meze české gramatiky*. Praha: Academia, 2006, s. 112–124. ISBN 80-200-1463-2.
- URREA, M. A. (2000). Automatic Discovery of Affixes by means of a Corpus: A Catalog of Spanish Affixes. *Journal of Quantitative Linguistics*, 2000, vol. 7, num. 2, pp. 97–114. ISSN 0929-6174.
- URREA, M. A. (2006). Affix Discovery by Means of Corpora: Experiments for Spanish, Czech, Ralámuli and Chuj. In *Aspects of Automatic Text Analysis*. Berlin: Springer, 2007, s. 227–299. ISBN 978-3-540-37520-3.
- URREA, M. A. – HLAVÁČOVÁ, J. (2005). Automatic Recognition of Czech Derivational Prefixes. In Gelbukh, A. (ed.). *CICLing 2005, LNCS*, vol. 3406. Berlin; Heidelberg: Springer, 2005, pp. 189–197. ISBN 978-3-540-24523-0.
- VÁCLAVÍK, J. (2007). Segmentace textových a biologických sekvencí metodou soutěživých vzorů. Prezentace k výuce na ÚFAL MFF UK. Dostupné na: <[http://is.muni.cz/th/3873/fi\\_m/dp01.pdf](http://is.muni.cz/th/3873/fi_m/dp01.pdf)> [cit. 9. 8. 2010].
- VACHEK, J. (2005). *Lingvistický slovník Pražské školy*. Praha: Karolinum, 2005. ISBN 80-246-0933-9.
- VEČERKA, R. (2004). Znepokojivá vágnost některých jazykovědných pojmů. In Karlík, P. – Pleskalová, J. (eds.). *Život s morfémy: Sborník studií na počest Zdenky Rusínové*. Brno: Masarykova univerzita v Brně, 2004, s. 165–178. ISBN 80-210-3373-8.
- VEČERKA, R. (2008). *Jazyky v komparaci I: Nástin české jazykovědné slavistiky v mezinárodním kontextu*. Praha: Nakladatelství Lidové noviny, 2008. ISBN 978-80-7106-575-3.
- VESELÝ, L. (2008). Práce o vidu založená na korpusu. *Naše řeč*, 2008, roč. 69, 4, s. 213–217. ISSN 0027-8203.
- VUŽŇÁKOVÁ, K. (2006). Slovník koreňových morfém slovenčiny a Malý slovotvorný slovník slovenčiny. In SOKOLOVÁ, M. – IVANOVÁ, M. – OLOŠTIAK, M. (eds.). *Morfematický výskum slovenčiny: Možnosti jeho štatistického elektronického a didaktického spracovanie*. Prešov: Filozofická fakulta Prešovskej univerzity v Prešove, 2006, s. 235–246. ISBN 80-8068-544-4.

- WEISHEITLOVÁ, J. (1981). Automatic Analysis of Czech Morphemics. *Prague Studies in Mathematical Linguistics*, 1981, [sv.] 7, s. 225–236.
- WEISHEITLOVÁ, J. – KRÁLÍKOVÁ, K. – SGALL, P. (1982). Morphemic Analysis of Czech. *Eksplícitnoje opisanije jazyka i avtomatičeskaja obrabotka tekstov = Explizite Beschreibung der Sprache und automatische Textbearbeitung = Explicitní popis jazyka a automatické zpracování textu*. Vol. VII. Praha: MFF UK, 1982. 120 s.
- WICENTOWSKI, R. (2004). Multilingual Noise-Robust Supervised Morphological Analysis using the WordFrame Model. In *ACL Workshops: Proceedings of the 7th Meeting of the ACL Special Interest Group in Computational Phonology: Current Themes in Computational Phonology and Morphology*. Barcelona, 2004, s. 70–77. Dostupné na: <<http://www.aclweb.org/anthology/W/W04/W04-0109.pdf>> [cit. 9. 8. 2010].
- WORTH, D. S. – KOZAK, A. S. – JOHNSON, D. B. (1970). *Russian derivational dictionary*. New York: American Elsevier, 1970.
- YUEN, R. – CHAN, T. – LAI, T. – KWONG –T'SOU, B. (2004). Morpheme-based Derivation of Bipolar Semantic Orientation of Chinese Words. In *International Conference On Computational Ling: Proceedings of the 20th international conference on Computational Linguistics*. Geneva, 2004. Dostupné na: <<http://www.aclweb.org/anthology/C/C04/C04-1145.pdf>> [cit. 9. 8. 2010].
- ZEMAN, D. (2007). Unsupervised Acquiring of Morphological Paradigms from Tokenized Text. In *Working Notes for the Cross Language Evaluation Forum (CLEF)*. Budapest, 2003. ISBN 2-912335-31-0. Dostupné na: <[http://www.clef-campaign.org/2007/working\\_notes/ZemanCLEF2007.pdf](http://www.clef-campaign.org/2007/working_notes/ZemanCLEF2007.pdf)> [cit. 9. 8. 2010].
- ZEMAN, D. (2009). Using Unsupervised Paradigm Acquisition for Prefixes. In *Evaluating Systems for Multilingual and Multimodal Information Access*. Berlin: Springer, 2009, s. 983–990. ISBN 978-3-642-04446-5. Dostupné na: <[http://clef-campaign.org/2008/working\\_notes/zeman-paperMC08.pdf](http://clef-campaign.org/2008/working_notes/zeman-paperMC08.pdf)> [cit. 9. 8. 2010].
- ZHANG, BYOUNG-TAK – KIM, YUNG-TAEK. (1990). Morphological Analysis and Synthesis by Automated Discovery and Acquisition of Linguistic Rules. In *Proceedings of the 13th conference on Computational linguistics*. Helsinki, 1990, s. 431–436.
- ZIKOVÁ, M. (2007). [Slovotvorba:] Počátek 21. Století. In Pleskalová, J., et al. (eds.). (2007). *Kapitoly z dějin české jazykovědné bohemistiky*. Praha: Academia, 2007, s.273–277. ISBN 978-80-200-1523-5.
- ZIKOVÁ, M. (2008). *Alternace vokálů s nulou v současné češtině – laterální autosegmentální analýza*. Brno, 2008. Disertační práce na FF MU. Dostupné na: <[http://is.muni.cz/th/9336/ff\\_d/](http://is.muni.cz/th/9336/ff_d/)> [cit. 9. 8. 2010].
- ZIKOVÁ, M. (2009). Ke vzniku fázových sufixů v češtině aneb Jak se domček změnil v domeček. *Slovo a slovesnost*, 2009, roč. 70, č. 4, s. 263–275. ISSN 0037-7031.

## Seznam obrázků a tabulek

### Seznam obrázků

Obr. 1: Ukázka Čelakovského kartoték (uspořádání podle kořenů).....	41
Obr. 2: Ukázka Čelakovského kartoték (uspořádání podle kořenů).....	42
Obr. 3: Ukázka Čelakovského kartoték (uspořádání podle zakončení).....	42
Obr. 4: Ukázka Čelakovského kartoték (sopis cizích slov).....	43
Obr. 5: Přehled rovin a jednotek (Sgall 1967: 56).....	45
Obr. 6: Přehled rovin a jednotek (Čermák 2001: 102n).....	46
Obr. 7: Model jazyka (Erhart 1973: 130).....	46
Obr. 8: Ukázka slovníku (RMSČ/Slavičková 1975: 412).....	64
Obr. 9: Ukázka seznamu kořenových morfémů (RMSČ/Slavičková 1975: 518).....	65
Obr. 10: Ukázka slovníku (Šiška 2005: 102).....	67
Obr. 11: Ukázka slovníku (Sokolová – Moško – Šimon – Benko 1999: 224).....	68
Obr. 12: Ukázka slovníku (Sokolová – Ološtiak – Ivanová 2005: 159).....	69
Obr. 13: Ukázka slovníku (Jadacka et al. 2001–2004, Tom 2, s. 955).....	70
Obr. 14: Ukázka slovníku (Worth – Kozak – Johnson 1970: 101).....	71
Obr. 15: Ukázka slovníku (Kuzněcova – Jefremova 1986: 281).....	72
Obr. 16: Ukázka slovníku (Tichonov 1996: 347).....	73
Obr. 17: Ukázka slovníku (Polikarpov 2000).....	73
Obr. 18: Ukázka slovníku (Poljuha 1983: 144).....	74
Obr. 19: Ukázka slovníku (Karpilovs'ka 2002: 377).....	75
Obr. 20: Ukázka slovníku (Augst 1975, vol. 1, s. 533).....	76
Obr. 21: Klasifikace segmentů v slovenštině (Sokolová 1999: 30).....	82

### Seznam tabulek

Tab. 1: Ukázka automatické morfematické segmentace (Cvrček 2009).....	85
Tab. 2: Struktura slovní zásoby podle číselné indexu z RMSČ.....	91
Tab. 3: Morfematická struktura češtiny podle počtu morfémů (celkový přehled).....	92
Tab. 4: Morfematická struktura češtiny podle počtu morfémů (a podle sl. druhů).....	93
Tab. 5: Morfematická struktura češtiny podle počtu slabik (celkový přehled).....	93
Tab. 6: Morfematická struktura češtiny podle počtu slabik a slovních druhů.....	94
Tab. 7: Morfematická struktura češtiny podle slovních druhů.....	94
Tab. 8: Morfematická struktura češtiny z hlediska typologie morfémů (podle pozice ve slově).....	94
Tab. 9: Morfematická struktura češtiny z hlediska typologie morfémů (podle pozice ve slově a sl. druhu).....	95
Tab. 10: Morfematická struktura češtiny z hlediska typologie morfémů (podle celé značky s vynecháním sl. druhu).....	96
Tab. 11: Morfematická struktura češtiny z hlediska typologie morfémů (podle celé značky).....	97
Tab. 12: 50 nejčastějších morfémů podle formy – grafematicky.....	98
Tab. 13: 50 nejčastějších morfémů podle formy a pozice ve slově – grafematicky.....	99
Tab. 14: 50 nejčastějších morfémů podle formy a celé značky – grafematicky.....	101
Tab. 15: 50 nejčastějších morfémů podle formy – fonematicky.....	102
Tab. 16: 50 nejčastějších morfémů podle formy a pozice ve slově – fonematicky.....	103
Tab. 17: 50 nejčastějších morfémů podle formy a celé značky – fonematicky.....	104
Tab. 18: 50 nejčastějších morfémů podle složení z konsonantů a vokálů – grafematicky (forma a celá značka).....	106
Tab. 19: 50 nejčastějších morfémů podle složení z konsonantů a vokálů – grafematicky (forma).....	107
Tab. 20: Morfém z hlediska počtu slabik.....	107
Tab. 21: 50 nejčastějších morfémů podle složení z konsonantů a vokálů – fonematicky (CSVD).....	108
Tab. 22: 50 nejčastějších morfémů podle složení z konsonantů a vokálů – fonematicky (CV: forma, celá značka).....	110

<i>Tab. 23: 50 nejčastějších morfémů podle složení z konsonantů a vokálů – fonematically (CV: forma) .....</i>	<i>111</i>
<i>Tab. 24: Vzorce struktury slov podle typů morfémů (preradix. část   radix. č.   postradix. č.) – řazení podle RMSČ.....</i>	<i>115</i>
<i>Tab. 25: Vzorce struktury slov podle typů morfémů (preradix. část   radix. č.   postradix. č.) – řazení podle RMSČ.....</i>	<i>120</i>
<i>Tab. 26: 50 nejčastějších vzorců struktury slov podle typů morfémů a slovního druhu (preradix. část   radix. č.   postradix. č.) – řazení podle RMSČ.....</i>	<i>121</i>
<i>Tab. 27: 50 nejčastějších vzorců struktury slov podle typů morfémů a slovního druhu (preradix. část   radix. č.   postradix. č.) – řazení podle synČNK.....</i>	<i>122</i>
<i>Tab. 28: 50 nejčastějších vzorců struktury slov podle celé značky s vynecháním slovního druhu (preradix. část   radix. č.   postradix. č.) – řazení podle RMSČ.....</i>	<i>123</i>
<i>Tab. 29: 50 nejčastějších vzorců struktury slov podle celé značky s vynecháním slovního druhu (preradix. část   radix. č.   postradix. č.) – řazení podle synČNK.....</i>	<i>125</i>
<i>Tab. 30: 50 nejčastějších vzorců struktury slov podle formy, celé značky a slovního druhu (preradix. část   postradix. č.) – řazení podle RMSČ.....</i>	<i>126</i>
<i>Tab. 31: 50 nejčastějších vzorců struktury slov podle formy, celé značky a slovního druhu (preradix. část   postradix. č.) – řazení podle synČNK.....</i>	<i>127</i>

# PŘÍLOHA

## 1.0 Původní znění návrhu projektu *Morfematická analýza češtiny*

Grantový návrh nakonec neschváleného projektu *Morfematická analýza češtiny* byl předložen Grantové agentuře Univerzity Karlovy v Praze 21. 11. 2006.

Český název projektu:	Morfematická analýza češtiny
Anglický název projektu:	Morphemic Analysis of Czech
První žadatel:	Jiří Lebeda
Studium:	Filozofická fakulta
Program:	Filologie
Obor:	Český jazyk a literatura
Typ studia:	magisterské studium
Rok založení projektu:	2007
Délka řešení projektu:	1
Sekce oborové rady:	Společenské vědy
Historie stavu:	<b>11. 10. 2006</b> – nový

51775757	Řešitel	Jiří Lebeda	stipendia: 48
-	Školitel	Vladimír Petkevič	mzdy: 14
[ID studenta č. 1]	Spoluřešitel	[jméno studenta č. 1]	stipendia: 24
[ID studenta č. 2]	Spoluřešitel	[jméno studenta č. 2]	stipendia: 24

### Charakteristika řešitelského kolektivu

Hlavní řešitel (J. Lebeda) bude v projektu zodpovědný za metodologickou stránku řešení, vypracuje algoritmus morfematické segmentace a bude se podílet na vyhodnocení jeho úspěšnosti. Spoluřešitelé ([student č. 1] a [student č. 2]) provedou finální korekturu a formální sjednocení rozsáhlých referenčních dat (dva morfematické slovníky a publikované seznamy tvarotvorných a slovotvorných jednotek), budou podle připravené metodiky anotovat větší část z výchozí báze o velikosti cca patnácti tisíc jedinečných morfémů a za pomoci korpusových rešerší uskuteční ověření navrženého algoritmu. Vedoucí práce (V. Petkevič) bude kromě poskytování odborných konzultací průběžně sledovat možnosti koordinovat práce na projektu s řešením výzkumného záměru „Český národní korpus a korpusy dalších jazyků“ a grantového projektu „Kapitoly z české gramatiky“, jejichž je spoluřešitelem, jakož i obou výzkumných záměrů Ústavu pro jazyk český AV ČR, kde působí jako člen vědecké rady. Cílem je, aby nedocházelo ke zbytečnému tříštění sil a prostředků a aby právě naopak bylo v tomto směru dosaženo synergického efektu.

### Zdůvodnění finančních požadavků projektu

#### 1) MZDY A STIPENDIA

Celkem: 115 tisíc korun

Rozsah i náročnost prací souvisejících s řešením projektu (viz ostatní oddíly této žádosti) kladou značné nároky na jeho personální a finanční zabezpečení. Předpokládané množství zpracovávaných jazykových dat je nezvládnutelné bez účasti obou spoluřešitelů a jejich ohodnocení v požadované výši.

#### 2) TECHNICKÉ VYBAVENÍ

Celkem: 60 tisíc korun

Z hlediska materiálového zázemí je pro charakter prací nezbytné disponovat dvěma přenosnými počítači (notebooky) vybavenými základním softwarem (operační systém a běžné kancelářské programy), jejichž pořizovací náklady jsou odhadovány souhrnně na 60 tisíc korun.

#### 3) SLUŽBY

Celkem: 65 tisíc korun

Zásadní podmínkou pro úspěšné řešení projektu je převod dvou tištěných morfematických slovníků a dalších meritorních pramenů do elektronické podoby (skenování a obtížné OCR zpracování cca 1200

stran textu), které bude zabezpečeno na základě nejvýhodnější nabídky dodavatelem tohoto druhu služeb. Odhad s tím spojených výdajů dosahuje 40 tisíc korun. Profesionální internetová prezentace výstupů projektu zahrnující též jednoduchou interaktivní demonstraci navrženého algoritmu, která může o dosažených výsledcích nejúčinněji informovat co nejširší okruh zájemců, představuje náklady ve výši 25 tisíc korun. Realizátor technického řešení prezentace bude rovněž vybrán formou výběrového řízení.

#### 4) OSTATNÍ

celkem: 6 tisíc korun

Rozpočet projektu počítá ještě se 3 tisíci korun na pokrytí výdajů spojených s nákupem potřebné odborné literatury a uhrazením konferenčních poplatků pro účast hlavního řešitele na tematicky blízkých konferencích konaných v průběhu řešení projektu. Tři tisíce korun vyčleněné na nespécifikované mzdové a stipendijní náklady představují rezervu, která bude použita na krytí „víceprací“ spoluřešitelů vynaložených na řešení projektu.

#### **Anotace**

Projekt morfemické analýzy češtiny usiluje o vytvoření algoritmu pro automatickou segmentaci českého slova na nejmenší významotvorné jednotky, tj. morfémy, založeného na pravidlech a podrobné klasifikaci typických segmentů včetně heuristik pro identifikaci jednotek nových a periferních (zejm. kořenů a afixoidů). Na základě dosavadních teoretických a empirických prací (slovníků a publikovaných seznamů) bude vypracována metodika aplikovatelná na libovolný český text (s jistým omezením pro propriální sféru jazyka), na ověření nosnosti navrženého postupu budou využity rozsáhlé jazykové korpusy psaného i mluveného jazyka spravované Ústavem Českého národního korpusu na Filozofické fakultě Univerzity Karlovy. Vůbec poprvé bude možné morfemicky analyzovat libovolný tvar slova (nikoli pouze lemma), segmentační algoritmus nebude omezen pouze na spisovnou varietu jazyka. Dnešní počítačové systémy pro morfologickou analýzu češtiny pracují se strojovými slovníky o velikosti řádově statisíců jednotek, pro účely úspěšné morfemické analýzy vystačíme řádově s tisíci jednotkami. Výsledky projektu otevrou cestu k efektivnímu výzkumu struktury současné češtiny na různých jazykových rovinách, umožní získat cenná statistická data a rozšíří dosavadní formální popis českého jazyka (zejm. morfologie, syntax) o nové poznatky (z morfonologie, morfematiky, slovotvorby) užitečné i pro účely výuky. Hotový algoritmus bude připraven pro budoucí počítačovou implementaci.

#### **Anotace v anglickém jazyce**

The project of morphemic analysis of Czech aims to develop the algorithm for automatic segmentation of Czech words into the smallest significant units, i.e. morphemes, based on the rules and detailed classification of typical segments including the heuristics for identification of unknown and peripheral units (mostly root morphemes and affixoids). On the basis of recent theoretical and empirical approaches we propose to evolve the methodology suitable for any input written in Czech. In order to verify the method proposed we will use corpora of spoken and written language. For the first time, morphemic analysis of any word form (not only lemma) will be possible, moreover, the algorithm will be applicable not only to literary language. Present-day automatic morphological analysers of Czech work with hundreds of thousand of items, for morphemic analysis about fifteen thousand will be sufficient. The results achieved in this project will allow more effective exploration of the Czech language structure on various strata and will facilitate acquisition of valuable statistical data as well. Existing formal description of Czech (primarily morphology and syntax) will be extended by new knowledge (from the area of morphonology, morphemics, word formation) which can further be used for education purposes. The resulting algorithm will be prepared for future computational implementation.

#### **Současný stav poznání**

Výzkum v oblasti morfematiky má ve světové i české lingvistice dlouhou tradici, a to jak na úrovni obecnělingvistické (de Courtenay, Saussure, Hjelmslev, Bloomfield, Harris, Gleason, Hockett), tak při rozpracování teorií pro konkrétní jazyky (Laskowski, Kuzněcova, Jefremova, Oliverius, Augst, Fleischer, Barz, Horecký, Bosák, Buzásová, Slavičková, Skoumalová, Šiška ad.). Velký význam má zkoumání morfemické struktury jazyka především pro jazyky flektivní, jakým je např. čeština,

pro něž je příznačná velká tvarová bohatost a variabilita (od několika set tisíc základních podob uváděných ve slovnících lze paradigmaticky utvořit odhadem až kolem osmi milionů tvarů). Není náhodné, že právě jazykovědná bohemistika a slovákistika dosáhly v oblasti strukturního popisu morfematické roviny jazyka metodologicky podnětných výsledků v teorii (Dokulil 1962, Slavičková 1962, Horecký 1964, Ďurovič 1965, Slavičková 1965, Daneš – Dokulil – Kuchař et al. 1967, Slavičková 1967, Horecký 1969, Skoumalová 1976, Bosák – Buzássyová 1985, Sokolová 1994, Horecký 1995, Sokolová – Šimon – Moško 1996, Sokolová – Ivanová – Vužňáková 2005 ad.) i praxi (Slavičková 1975, Sokolová et al. 1999, Šiška 2005, Sokolová – Ološtiak – Ivanová 2005).

Slovenská jazykověda v posledních letech postoupila na jedné straně od studia morfematiky k výzkumu slootovorb (Vužňáková 2006), na straně druhé se orientuje též na korelace morfémů s jednotkami nižších rovin, např. slabikou (Sabol – Bónová 2003). Česká lingvistika se v posledních desetiletích s výjimkou edukativně orientované práce Šiškovy o morfematiku prakticky přestala zajímat, slootovorbě je věnována částečná pozornost v rámci přípravy empiricky orientované gramatiky současné češtiny založené na korpusovém materiálu (F. Štícha a kolektiv). Nevýhodou dosavadních přístupů k morfematické analýze a na nich založených metod segmentace je jejich manuální, neautomatizovaná podstata a s tím související omezený rozsah takto zpracovaných dat. Morfematické slovníky pro češtinu a slovenštinu vycházejí z lexikální báze o velikosti 60–70 tisíc lemmat. Hendikepem je rovněž (do značné míry odůvodnitelná) redukce zkoumaného materiálu na základní tvary slov a spisovnou varietu jazyka. Širšímu využití dosažených výsledků brání též zpracování pouze grafické podoby analyzovaných slov bez přihlížení ke stavbě fonologické. Zajímavé výsledky by jistě přinesla konfrontace morfematických a slabičných švů, která rovněž zatím nebyla realizována.

Zdálo by se, že uvedené nedostatky mohou být eliminovány se vznikem a praktickým nasazením automatických programových nástrojů pro analýzu češtiny, které jsou vytvářeny a postupně zdokonalovány od konce osmdesátých let. Skutečnost však není natolik příznivá, jak by se mohlo na první pohled zdát. Pro morfologickou analýzu a syntézu (včetně kontroly pravopisu, tzv. spelling checker) je v českém prostředí užíváno zhruba těchto pět šest programů: dva akademické – pražský tzv. Hajičův (viz Hajič 2004) a brněnský nesoucí název Ajka (Sedláček 2004) –, „svobodný“ program Ispell (resp. jemu blízký Aspell) – a dva komerční – Lemma nabízený brněnskou firmou Lingea a nepojmenovaný analyzátor distribuovaný firmou Eridanus. Žádný z těchto nástrojů není postaven na čistě lingvistických základech a ani jeden z nich neposkytuje uspokojivé výsledky v oblasti morfematické segmentace. Algoritmické pozadí těchto programů se často uchyluje k technickému řešení formou seznamu „idiosynkratických“ výjimek či zbytkové třídě tzv. intersegmentů. Ajka je dostupná také ve verzi analyzující – díky ručně realizovanému propojení některých tvarotvorných a slootovorných vzorů – vybrané slootovorné vztahy, pro tzv. Hajičův morfologický analyzátor existují pokusy o podpůrný nástroj pro „hádání“ parametrů neznámých slov (tzv. guesser – Hlaváčová, Votrubec). Lingvisticky motivovaný algoritmus morfematické segmentace, který je cílem tohoto projektu, skýtá – v případě efektivní implementace – nemalý potenciál pro rozvoj automatických analyzátorů češtiny směrem, který je za současné situace přístupný jen omezeně (analýza slootovorb v plném rozsahu, snazší identifikace jazykových variant, zpracování neznámých slov, možnost odhalení některých typů tzv. lexikální homonymie apod.).

#### Literatura:

[vynechán seznam literatury o 22 bibliografických záznamech tvořící podmnožinu soupisu pramenů této práce]

#### **Přínos projektu k rozvoji fakulty / VŠ**

Předkládaný projekt se svým zaměřením na funkčně-strukturní analýzu morfematického systému češtiny hlásí k živému odkazu Pražského lingvistického kroužku a pražské lingvistické školy spjatému také s Filozofickou fakultou Univerzity Karlovy. Jeho metodologie i praktické výsledky se mohou stát účinnou inspirací pro ostatní, především slovanské, filologické obory, které jsou na fakultě pěstovány. Svým vymezením má projekt blízko ke dvěma výzkumným záměrům řešeným na Filozofické fakultě UK, a to 1) Český národní korpus a korpusy dalších jazyků (MSM0021620823), hlavní řešitel F. Čermák (ÚČNK) a 2) Jazyk jako lidská činnost, její produkt a faktor (MSM0021620825), hlavní řešitel Z. Starý (ÚLUG). Obsahově je projekt rovněž v souladu s Aktualizovanými Dlouhodobými

základními směry výzkumu schválenými vládou České republiky 18. 10. 2006, a to především s kapitolami Informační společnost a Společenskovědní výzkum.

### **Materiální zajištění projektu**

Projekt bude materiálově a přístrojově zajištěn především dvěma notebooky včetně nezbytného programového vybavení (hlavní řešitel a oba spoluřešitelé) pořízených z prostředků grantu, zpracovávaná data budou získána prostřednictvím objednaných služeb (viz Zdůvodnění finančních požadavků projektu a Způsob řešení) a díky rešerším zabezpečeným spoluřešiteli. Vedoucí práce bude využívat zázemí svého mateřského pracoviště (ÚTKL FF).

### **Cíle řešení projektu**

Česká republika definuje svou jedinečnost v rámci integrující se Evropy i na pozadí globalizujícího se světa především prostřednictvím jazyka, historie a kultury, přičemž právě jazyk je hodnotou nejvýznamnější, neboť dějiny i kultura jsou zprostředkovávány zpravidla jeho prostřednictvím. Čeština je pro obyvatele České republiky nejen jazykem mateřským, ale je rovněž jedním z dvaceti úředních a jednacích jazyků Evropské unie. Na území EU existuje přibližně osmdesát původních evropských jazyků představujících kulturní a jazykové dědictví nesmírné hodnoty, které je nutné chránit a dále rozvíjet. Zájem o český jazyk nevzrůstá výlučně z potřeb utilitárních (překladatelské služby, přeshraniční spolupráce), nýbrž jde dnes ruku v ruce se zvyšujícími se nároky informační společnosti na dostupnost, shromažďování a zpracování enormního množství informací vyjádřených především pomocí psaného jazyka. Vyhledávání informací (vžívá se též anglický termín „informational retrieval“) je dnes jedním z nejčastějších způsobů využití internetu, záplava dat však vyžaduje vývoj stále nových a nových celosvětově použitelných metod a technologií, které budou schopné vyrovnat se s jazykovými rozdíly na straně poskytovatele a příjemce obsahu (např. strojový překlad, dialogové systémy v přirozeném jazyce apod.). Podpora češtiny v jednotlivých aplikacích (v celosvětových i národních vyhledávacích typu Google, v elektronických katalozích knihoven a archivů, na portálech státních institucí, v projektech e-learningu aj.) zatím stále není optimální. Základním předpokladem pokroku v této věci je co nejzevrubnější formální popis přirozeného jazyka, v našem případě češtiny.

Proponovaný cíl tohoto projektu vychází vstříc jak potřebám teoretického popisu, tak nárokům aplikační sféry. Algoritmus pro morfematickou segmentaci češtiny, který je hlavním výstupem řešení tohoto projektu, představuje skutečně široké možnosti využití. Jeho prostřednictvím je např. možné dosáhnout hlubšího poznání stavby českého jazyka na morfologické a hierarchicky nižších rovinách (zkoumání kombinatoriky morfémů, stanovení inventáře kořenů českého lexika aj.) včetně rozpoznání vývojových tendencí v přejímání a adaptaci slov cizího původu (neologismů). Nabízí se tak rovněž účinný nástroj pro výzkum české slovo tvorby a variantnosti formální morfologie. Tím se v neposlední řadě otevírá prostor pro principiálně nové metody výuky češtiny založené na práci s živým, autentickým jazykovým materiálem, neboť morfematická analýza může pracovat s libovolným vstupním textem. Z hlediska aplikačního algoritmus jednak skýtá značný potenciál pro zdokonalování existujících nástrojů pro zpracování češtiny (guesser, rozšiřování pokrytí slovní zásoby, derivační vztahy atd.), jednak podněcuje vznik nástrojů nových (např. pro uložení a komprimaci rozsáhlých dat založené na opakovatelnosti morfémů jako nejnižšího počtu funkčních jednotek v jazyce). Zhodnocení výsledků je závislé také na tom, jaká péče bude po vyřešení tohoto grantu věnována dalšímu kvantitativnímu i kvalitativnímu rozvoji segmentačních pravidel, doplňování dosud neregistrovaných klasifikovaných segmentů a vývoji nových funkcí (jemnější analýza proprií, slabičné dělení, morfematická syntéza atp.).

### **Způsob řešení**

Projekt lze podle charakteru prací označit jako empirický, aplikační, svým způsobem půjde o syntézu a doplnění dosavadních poznatků, které budou experimentálně ověřovány na surovém (tj. autentickém, nepředzpracovaném) referenčním datovém materiálu. Nejprve bude z hlediska možností formalizace dosavadních metod morfematické segmentace věnována pozornost problematikým jevům, v jejichž hodnocení a způsobech řešení nepanuje v literatuře shoda, resp. budou rozlišeny kategorie, které nelze rozhodovat na pravidlovém základě (např. deetymologizace: po(-)šet-i-l-ý; reziduální morfémy: mal-in-a, protože ostruž-in-a; překrývání morfémů: knih-kup-e-[c]-ctv-í, ru[s]-ský aj.). Vyjasněna bude



rovněž problematika diachronního a synchronního přístupu k segmentaci i otázka uchopení fenoménu nulového morfu, resp. submorfu (den-0, d0n-e, z-krať-0-t-e, z-krať-0-0, kuř-0-e). Paralelně s definováním efektivní metody automatické (tj. pravidly řízené) segmentace budou probíhat práce na digitalizaci a korektuře referenčních dat (Retrogradní morfematický slovník češtiny E. Slavíčkové, Šiškův Bázový morfematický slovník češtiny, inventáře slovotvorných a tvarotvorných formantů z Tvoření slov v češtině 2 a dalších pramenů), která budou dále podrobně analyzována a klasifikována zejm. z hlediska významu – na morfémy (kořeny, derivační, modifikační a gramatické morfémy) a submorfy (infixy/konektémy a interfixy) – a podle kombinatorických vlastností – na volné a vázané, rozštěpené, prefixální, sufixální, postfixální atd.

Pro identifikaci kořenů bude vyvinut speciální postup založený na typologizaci kořených morfů (vyčleněných v referenčních zdrojích), jenž bude podpůrně využívat pravidlovou fonologickou transkripci. Je to prakticky jediný – dosud téměř nevyužívaný – způsob, jak se lze vyrovnat s výrazovou (grafematickou) variabilitou (/skɛnr/: scanner, scaner, skener, skenr...), zároveň tak máme možnost evidovat vokální a konsonantické alternace projevující se v jiných než základních tvarech (/po-VLAC-ích, po-vlak-0, po-vleč-e-ň-í, vlek-0-l-í, na-vlák-a-t, vláč-e-t, vléc-0-i/). Heuristickým způsobem bude zpracována problematika segmentační homonymie, a to jak homonymie formální (pro(-)vaz × vaz-0), tak i systémové (tříd-i-č-k-a × tříd-ič-k-a). Nosnost navržené metodiky bude ověřována na datech textových korpusů spravovaných Ústavem Českého národního korpusu na FF UK, úspěšnost segmentace bude testována na ručně zpracovaném materiálu (morfematických slovnících) a vyhodnocena pomocí standardních hodnot přesnosti (precision) a úplnosti (recall). Hotový algoritmus bude mít podobu zpracované soustavy segmentačních pravidel (např. o komplementární distribuci různých typů segmentů se stejnou funkcí) a anotovanýchází morfémů.

## **Prezentace výsledků**

Výsledky budou prezentovány prostřednictvím nejméně dvou článků publikovaných v odborném lingvistickém časopise, plánován je přinejmenším jeden přednesený příspěvek na jazykovědné konferenci (zvažováno je vystoupení na 2. mezinárodní konferenci „Gramatika a korpus“ pořádané Ústavem pro jazyk český AV ČR v září 2007 spojené navíc s publikováním abstraktu referátu). Dalším chystaným výstupem je diplomová práce hlavního řešitele, která bude široké odborné i laické veřejnosti zpřístupněna prostřednictvím internetové prezentace. Pro tuto prezentaci počítáme také s jednoduchou interaktivní demonstrací navrženého algoritmu. Praktické výsledky projektu budou nabídnuty pro účely výuky v seminářích vyučovaných na FF Ústavem českého jazyka a teorie komunikace (zejm. Proseminář lingvistické bohemistiky, Lexikologie a slovotvorba a Morfologický seminář).

## **1.1 Posudek č. 1 [ztajeným autorem zřejmě J. Šimandl]**

*Vědecká závažnost a aktuálnost projektu*

V projektu jde sice o zpracování poznatků dostupných v knižní literatuře z větší části i na konkrétním materiálu, z nezanedbatelné části však jen na úrovni teorie a zásad. Bude třeba je systematizovat a promítnout do jazykového materiálu in extenso, což zatím nikde není provedeno v univerzálně použitelné kvalitě, A HLAVNĚ zpracovat je nově v elektronické podobě a vytvořit tak předpoklady jejich aplikačního využití ve stále se rozrůstající sféře počítačového zpracování češtiny.

*Zpracování návrhu projektu a reálnost cílů řešení*

Projekt hodnotím jako jasně formulovaný a proveditelný.

*Koncepce a metodika*

Metodický přístup pokládám za adekvátní zpracovávané problematice i cílům.

*Vybavení pracoviště a časový plán*

Splnění cílů je reálné očekávat a pracoviště doc. Petkeviče představuje velmi kredibilního garanta.

*Řešitelský kolektiv*

Publikační i pedagogická aktivita školitele, jeho participace na národně, evropsky i světově významných projektech (Český národní korpus, jazyková pracoviště MFF UK) dává v rámci FF snad nejvyšší možné záruky. K ŘEŠITELI poznamenávám, že během svých studií pracoval na počítačném zpracování češtiny i pro Ústav pro jazyk český AV ČR, a to s velmi dobrými výsledky.

*Finanční náklady*

Rozpočet projektu je rozhodně přiměřený, spíš chvályhodně skromný.

### *Celkový komentář k návrhu projektu*

Kromě toho, co jsem uvedl v bodu 1, chci upozornit na to, že projekt vytváří krajně potřebnou konkurenci k nejrůznějším zpracováním téže problematiky od matematiků naprosto nedisponovaných studiím bohemistiky, u nichž se projevuje tendence k technicistním, příležitostným, účelovým, lingvisticky neobhajitelným řešením.

## **1.2 Posudek č. 2 [neodtajněným autorem nepochybně K. Pala]**

### *Vědecká závažnost a aktuálnost projektu*

Projekt zcela jistě směřuje k prohloubení současných znalostí v oboru tvarosloví češtiny. Segmentaci slov na "minimální" morfémy dosud nikdo nezpracovával, v tomto směru je tedy projekt originální. Je ovšem otázkou, zda projekt je směřován správným směrem - v době, kdy existuje úplný morfologický slovník češtiny pro elektronické zpracování, je cíl projektu "zmenšit stávající slovník (v elektronické formě)" poněkud redundantní, ačkoli vlastní problém segmentace relevantní je (např. pro lexikologii jako takovou).

### *Zpracování návrhu projektu a reálnost cílů řešení*

Cíle projektu jsou deklarovány jako teoretické i praktické. Cíle teoretické jsou více než legitimní; jak jsem uvedl výše, pro češtinu neexistuje dosud žádná automatická metoda plné segmentace slova na morfémy. Praktická část je ovšem značně nejasná: takový algoritmus nepřinese pro analýzu na tvaroslovné úrovni v zásadě mnoho nového, neboť při analýze jde především o identifikaci celého lexému (a zde již automatické nástroje existují) a morfématická segmentace zde nemá prakticky žádnou důležitost mimo morfologickou analýzu samotnou. V tomto směru by bylo vhodnější soustředit se na generování (syntézu), která by byla využitelná například při překladu (autoři návrhu správně zmiňují např. problém slovotvorby, který není v současných nástrojích plně vyřešen; nicméně ani zde se nezmiňují o tom, že největším problémem je iregularita, nikoli segmentace jako taková). Ačkoli to v odstavci o cílech projektu není uvedeno, cílem je zřejmě i zpřístupnit naskenovaný morfématický(é) slovník(y) (plánuje se na to poměrně podstatné množství finančních prostředků), což je samo o sobě jistě cíl záslužný.

### *Koncepce a metodika*

Koncepce a metodika projektu je nepříliš ujasněná. Skenování dvou současných zdrojů morfématické segmentace je vhodným předpokladem, ale není jasné, zda příslušná externí firma zajistí i kontroly výsledného OCR (zřejmě nikoli, v projektu se mluví o korekturách řešiteli), tj. převedení slovníků do spolehlivé textové podoby – bez ní však nebude možno tyto zdroje použít pro podporu cílů projektu, například finální evaluaci vytvořených algoritmů (která je jinak navržena vhodně a správně). Přes uvedení příkladů není jasné, na jakém základě budou vytvořené algoritmy implementovány – metodami formálních automatů a gramatik nebo „prostým programováním“? Na druhou stranu je nutno poznamenat, že se jedná o výzkum s mnoha neznámými, a že metodika bude jistě upravena podle postupu prací.

### *Vybavení pracoviště a časový plán*

Vybavení pracoviště je pro daný projekt dostatečné a nemám k němu připomínky. Časový postup není v plánu uveden, pro jeden rok však není příliš podstatný; podle mého názoru je však časový plán výrazně podhodnocen: i když bude skenování a OCR zadáno externí firmě, jsou jen korektury OCR a základní zpracování do použitelné podoby (pro jakýkoli účel!) značně časově náročné a spolehlivé převedení včetně kontrol, konzistence apod. zabere nejméně jeden rok pro dva až tři pracovníky. Vlastní vědecké úkoly jsou rovněž značně rozsáhlé, a to i kdyby řešitelé pracovali (jak se mi zdá rozumné) spíše na teoretické části projektu, tj. přípravě formálního rámce, rozlišení typů segmentů, přístup a zpracování nulových segmentů, vyjasnění diachronních diferencí apod. To samo o sobě je časově náročné (nutnost ověřování na získaném materiálu), byť pochopitelně velmi legitimní z hlediska postupného dosažení cíle(ů).

### *Řešitelský kolektiv*

Práci řešitelského kolektivu má koordinovat doc. V. Petkevič, jehož vynikající kvalifikace v oblasti tvarosloví je mimo vši pochybnost. Zde nemám absolutně žádné připomínky.

### *Finanční náklady*

Rozpočet projektů GAUK je obtížné posuzovat, neboť z principu věci musí disponovat dalšími zdroji. Vzhledem ke stanoveným cílům je rozpočet podle mého názoru podhodnocený (viz odstavec o

časovém plánu projektu); zde je však na navrhovatelích, aby posoudili své zdroje sami – je evidentní, že další podporu mají (např. VZ).

#### *Celkový komentář k návrhu projektu*

Celkově je tematika projektu z hlediska zpracování přirozeného jazyka důležitá a práce na daném tématu nade vše pochybnost nutná, zejména v kontextu EU, kde češtinu je nutno zpracovávat v ČR. Synergistický efekt projektu je rovněž dobře definován z hlediska tvorby nového slovníku v rámci VZ a spolupráce s ÚJČ AV ČR. Nicméně pro úspěšné splnění cílů projektu bych doporučoval dvě podstatné modifikace: prodloužení projektu na 3 roky (je podle mne plně odůvodněno šíří cílů projektu) a zároveň zúžení cíle na teoretickou část navržených cílů, tj. přesnou specifikaci morfématické segmentace a anotace (včetně vazeb na existující elektronické zdroje, jako jsou morfologické slovníky a slovníky syntaktické a sémantické (valenční)) a její ověření na daném, existujícím slovníkovém materiálu (po jeho převedení do elektronické podoby – tato část by měla v projektu nutně zůstat) a konverzi tohoto materiálu podle navržené specifikace. Procedury automatického zpracování češtiny z hlediska morfologické segmentace by si pak zasloužilo separátní projekt, ze zaměřením spíše na automatickou syntézu textu v takovém kontextu, kde nejsou k dispozici lemata (neboť pro zmámá lemata je tento problém je již úspěšně vyřešen). Do výstupů projektu je nutno v každém případě zařadit i zveřejnění elektronické verze naskenovaných zdrojů alespoň pro použití odbornou veřejností (CD, ke stažení apod. – ne jen internetový přístup, byť s vyhledáváním).